International Series in
Operations Research & Management Science

Cengiz Kahraman
Y. Ilker Topcu *Editors*

Operations Research Applications in Health Care Management

Springer

International Series in Operations Research & Management Science

Volume 262

More information about this series at http://www.springer.com/series/6161

Cengiz Kahraman • Y. Ilker Topcu
Editors

Operations Research Applications in Health Care Management

Springer

Editors
Cengiz Kahraman
Department of Industrial Engineering
Istanbul Technical University
Macka, Istanbul, Turkey

Y. Ilker Topcu
Department of Industrial Engineering
Istanbul Technical University
Macka, Istanbul, Turkey

ISSN 0884-8289 ISSN 2214-7934 (electronic)
International Series in Operations Research & Management Science
ISBN 978-3-319-65453-9 ISBN 978-3-319-65455-3 (eBook)
https://doi.org/10.1007/978-3-319-65455-3

Library of Congress Control Number: 2017959144

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

I dedicate this book to my Professors:

Prof. Coşkun Külür,
Prof. Necmettin Ongar,
Prof. Ataç Soysal,
Prof. Ahmet Fahri Özok,
Prof. Ethem Tolga,
Prof. Nahit Serarslan,
Prof. Haluk Erkut.

Prof. Cengiz Kahraman

I dedicate this book to
my late grandparents
who had been my inspiration and motivation
for continuing to improve my knowledge and
move my career forward

Prof. Y. Ilker Topcu

Preface

Healthcare management is composed of management activities and functions involved in various departments of healthcare systems. The objective of healthcare management is to influence the growth, development, and operations of healthcare systems. In recent years, there has been a significant increase in the interest for designing healthcare systems in order to address complex healthcare problems.

Operations research (OR) aims at developing mathematical and computational support for the optimization of problems in industry, service, and business. OR techniques have been used in the solutions of various healthcare problems in the literature. These studies have been published in some journals, proceedings, and books whose focus area is not totally healthcare systems. The motivation for editing this book has been the need for collecting the OR techniques applied to healthcare systems in a single source.

Twenty-two chapters have been submitted from various countries, namely, Turkey, the UK, the USA, Belgium, Croatia, Portugal, the Netherlands, Canada, Iran, Singapore, and Italy. This book is composed of eight main sections and two or three chapters under each main section.

The first chapter presents the taxonomy of operations research methodology in healthcare management to provide a common terminology and a classification mechanism.

The second chapter summarizes quantitative and qualitative techniques used in healthcare management, including OR techniques, statistical techniques, multicriteria decision making techniques, and others and presents graphical analyses of the survey results.

The third chapter proposes online optimization approaches for real time management of operating rooms. Real time management is capable to deal with the elective and non-elective patient flows within a single surgical pathway and with the resource sharing among different surgical pathways of elective patients. The authors assess the effectiveness of the proposed solutions on simulated surgical clinical pathways under several scenarios.

The fourth chapter presents two novel approaches for the identification of Takagi-Sugeno fuzzy models with time variant and time invariant features. The mixed fuzzy

clustering algorithm is used for determining the parameters of Takagi-Sugeno fuzzy models in two different ways. The fuzzy modeling approaches are tested on four healthcare applications for the classification of critically ill patients.

The fifth chapter reviews OR literature applied to hospital wards. The authors distinguish intensive care, acute medical units, obstetric wards, weekday wards, and general wards. They aim at guiding both researchers and healthcare professionals in identifying which OR technique/model suits best for each specific hospital ward setting.

The sixth chapter monitors the impact of interrelations in the development of an efficient and proactive system of chronic care management through Social Network Analysis.

The seventh chapter evaluates healthcare system efficiency of 34 OECD member countries using Data Envelopment Analysis (DEA). The base model is an output-oriented Banker-Charnes-Cooper model that uses the number of physicians, nurses, beds per 1000 population as inputs, and life expectancy at birth, infant survival rate as outputs.

The eighth chapter aims at predicting the healthcare expenditure per capita. Accurate estimation of healthcare expenditure can guide efficient healthcare policy making and resource allocation. Three different strategies are proposed to improve the forecasting accuracy of gray forecasting models. Genetic algorithm is applied for training data size and parameter optimization.

The ninth chapter focuses on vaccination and investigates the vaccine supply chains. This chapter aims at classifying some problems of the vaccine supply management which can be solved by mathematical programming tools.

The tenth chapter discusses the challenges and research opportunities in the blood collection operations and explores the benefits of recent advances in the blood donation process.

The eleventh chapter aims at introducing the recent developments in organ transplantation network planning as well as presenting relevant case studies. It focuses on mathematical programming and computational models proposed in the recent literature for organ transplantation network planning.

The twelfth chapter investigates a fuzzy decision tree algorithm applied to the classification of gene expression data. The fuzzy decision tree algorithm is compared to a classical decision tree algorithm as well as other well-known data mining algorithms commonly applied to classification tasks.

The thirteenth chapter presents an overview of disease screening problems and operations research applications on different aspects of the problems. The authors first discuss operations research applications in evaluation and optimization of screening policies and then organization of screening services for reaching out to the population and improving the effectiveness of screening.

The fourteenth chapter presents an analysis of the efficiency of diabetes treatment in the UK healthcare facilities using TOPSIS and neural networks. The authors provide a rational framework for policy makers to rank the efficiency of diabetes care facilities and also highlight the most important contextual variables that impact on the efficiency as issues of interest for future policies.

The fifteenth chapter uses achievement scalarization to obtain efficient solutions for radiation treatment planning. The authors adapt the parameters of the achievement scalarization to address a solution in a rectangle that is defined by the bounds on the objective functions. They compare the proposed method with multiobjective solution algorithm from the literature and clinical plans.

The sixteenth chapter introduces and discusses the recent developments of OR techniques for emergency medicals service management. Two selected mathematical models from the relevant literature are also elaborated. In addition, a real emergency medicals service location problem is described as a case study.

The seventeenth chapter is on medical informatics. A review of medical informatics is presented and a multidisciplinary point of view is given based on different approaches.

The eighteenth chapter aims at identifying the prevalent challenges of pharmaceutical supply chains at three different decision levels, i.e., long-term (strategic), mid-term (tactical), and short-term (operational) decisions.

The nineteenth chapter proposes a categorical data envelopment analysis framework for evaluating medical tourism performance of top destinations. Research hypotheses are created to analyze the relationship between the countries' medical tourism performance and their political, regulatory environment, technology, and knowledge outputs.

The twentieth chapter discusses different analytical techniques used in healthcare human resource planning. Two case studies are presented to provide examples of real-world applications across different institutional contexts.

The twenty-first chapter discusses the lean management techniques, their applications in the healthcare systems and how they can improve the performance of these systems by providing better patient service, better utilization of assets, and better patient care.

The twenty-second chapter introduces the procurement management in healthcare systems and discusses the related challenges together with some optimization approaches on procurement management problems in healthcare systems.

This book will provide a useful resource of ideas, techniques, and methods for the research on the theory and applications of OR techniques in healthcare management. We thank all the authors whose contributions and efforts made the publication of this book possible. We are grateful to the referees for their invaluable and highly appreciated works contributed to select the high quality of chapters published in this book.

Macka, Istanbul, Turkey

Cengiz Kahraman
Y. Ilker Topcu

Contents

About the Editors

Cengiz Kahraman is a full professor at Istanbul Technical University. His research areas are engineering economics, quality management, statistical decision making, multicriteria decision making, and fuzzy decision making. He has published more than 200 journal papers, about 150 conference papers, and 80 book chapters. He has guest-edited some issues of many international journals. He is the editor of many international books from Springer and Atlantis Press. He is the member of editorial boards of 20 international journals. He was the vice dean of ITU Management Faculty between 2004 and 2007 and the head of ITU Industrial Engineering Department between 2010 and 2013.

Y. Ilker Topcu is a Professor of Decision Sciences at Industrial Engineering Department of Istanbul Technical University (ITU) Management Faculty. He completed his Ph.D. studies in Engineering Management (2000) at ITU and visited Leeds University Business School during these studies (1998–1999). Professor Topcu's research interests include multiple criteria decision making, decision analysis, and operations research/management science. He is an executive committee member of International Society of Multiple Criteria Decision Making. He has published in several journals.

Part I
Overview

Chapter 1
A Taxonomy of Operations Research Studies in Healthcare Management

Serhat Tüzün and Y. Ilker Topcu

1.1 Introduction

The goal of researchers working in healthcare management is to control the rising costs and to increase accessibility for healthcare services. They try to do this by integrating the aspects of management with Operations Research (OR) techniques to determine the most efficient (or optimal) methods of providing patient care delivery (Langabeer 2007). The studies of OR in healthcare are not only about determining the methods for healthcare delivery, but also about simulating clinical systems to observe long-term risks.

Operations Research reached the stage of maturity in a very short time after it was first applied during WWII (Kirby 2003). It has been considered a discipline hard to grasp even though it spread to a wide application area. OR techniques are used to model and solve real-world problems in different areas such as production, logistics, etc. (Hillier and Lieberman 2005). Healthcare is yet another area, a relatively new one that Operations Research techniques are used in.

Healthcare is a business-like no other. Carter (2002) pointed out that it has multiple decision-makers with conflicting goals and objectives. Moreover, healthcare business has a high level of uncertainty as well as dynamic relationships among its components. Also, the managers in healthcare demand to lower the costs and increase the service quality. These aspects render management of healthcare and its operations reasonable to be studied with OR techniques.

S. Tüzün (✉)
Department of Industrial Engineering, Yildiz Technical University, Istanbul, Turkey
e-mail: serhat.tuzun@gmail.com

Y.I. Topcu
Department of Industrial Engineering, Istanbul Technical University, Macka, Istanbul, Turkey

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_1

The objectives of OR studies in Healthcare Management are to control the costs and to improve the quality of healthcare services (Mclaughlin and Hays 2008). For the last two decades, hundreds of articles were published, special journal issues were put together, and conferences were organized. Various studies are carried out in different areas of healthcare. Some of these areas are resource allocation, scheduling, managing waiting lists, streamlining patient flows, facility location, cost-effectiveness analysis, triage in emergency services and disease treatment investigations (Pierskalla and Brailer 1994).

Although a comprehensive taxonomic classification was made by Hulshof et al. (2012), an up-to-date taxonomy is still a necessity. This chapter aims to provide a general overview of OR studies in healthcare management, using a different taxonomy approach than that of Hulshof et al. The literature has been thoroughly reviewed, and by classifying previous studies according to their preferences, a taxonomy for OR studies in healthcare management has been prepared.

1.2 OR Studies in Healthcare Management

Healthcare management research was first established in the 1930s. Although there had been some studies before, application of operations research in healthcare is accepted to have started during the 1970s. First publications were mainly about health planning and administration (Stimson and Stimson 1972; Shuman et al. 1975; Fries 1981). Later on, research areas on healthcare widely spread from top management to the smallest operation.

OR studies in healthcare management became popular during the 1990s. Using OR techniques in healthcare attracted a lot of attention in many countries and lots of studies are currently carried out (Luss and Rosenwein 1997). Many universities and research groups have shown interest in the subject. For example, McGill School of Environment (MSE) has a program called "Healthcare Operations & Information Management," directed by Vedat Verter. Although research is generally centered in the USA and Canada, a working group of Association of European Operational Research Societies (EURO) called Operational Research Applied to Health Services (ORAHS), which was initiated in 1975, provides a network for researchers involved in the application of systematic and quantitative analysis in support of planning and management in the health services sector.

There are already some bibliographic studies that organize the papers and classify them (Flagle 1962) classified the problems encountered in the area. Fries (1976) organized the papers before 1975; and the literature between 1970 and 1989 was classified by Corner and Kirkwood (1991). Also, minor classifications were made in the following years. Preater prepared a bibliography on the application of queuing theory in healthcare and medicine (2002); Cayirli and Veral reviewed the literature of outpatient scheduling in healthcare (2003); Lowery (1996) and Jun et al. (1999) investigated the simulation applications in health services. Due to these bibliographic studies, this study excludes papers published before 2000, and focuses on more recent years where the literature is building up more quickly than ever.

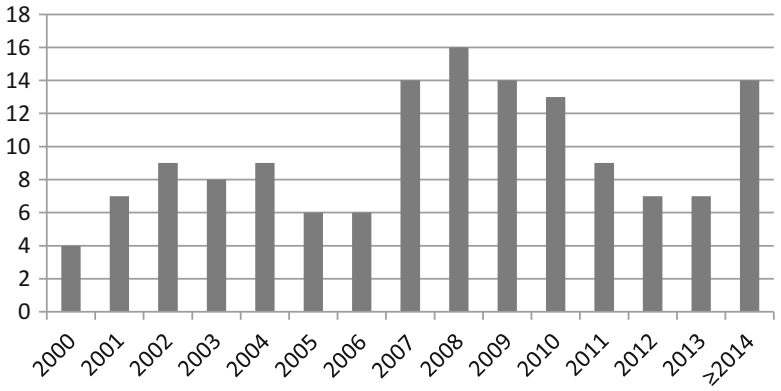

Fig. 1.1 Number of publications by years

The following keywords were used to find research papers published in the literature: healthcare operations management, healthcare management, health services, healthcare applications, health workforce planning, ambulance allocation, hospital resource allocation, outpatient scheduling, nurse–patient assignment, healthcare delivery, doctor/nurse workload, operating room planning, healthcare operations, doctor/nurse scheduling, health care production, emergency patient flow, health care services, and management decision support in the health service. The results of the search yielded over 500 articles, mainly in these journals: Annals of Operations Research, Artificial Intelligence in Medicine, Computers & Operations Research, European Journal of Operational Research, Expert Systems with Applications, Health Care Management Science, Health Policy, IIE Transactions, Interfaces, Omega, Social Science & Medicine, and Socio-Economic Planning Sciences. Before proceeding to the taxonomy, some exclusion criteria were determined in order to narrow the findings. Thus, we ended up with articles that were more related to the subject of "application of OR to healthcare management". These exclusion criteria were studies not in English, studies without models (Review papers), studies about improving treatment and diagnosis (screening, analyzing outputs, etc.), models based on probability and statistics, and models based on economic theory. As a result, 142 articles were within the criteria. Their distribution by the year of publication may be seen in Fig. 1.1.

1.3 Necessity for a Taxonomy: A Discussion

The size and growth rate of the literature demands a systematic way to classify various contributions in a manner that will vividly provide a panoramic view of what exists and will also clearly identify any existing gaps as suggested by Reisman (1992) and Reisman (1993).

Taxonomy may be defined as the science of identifying objects, and arranging them in a classification. According to Gattoufi et al. (2009), it is not only an efficient and effective tool for systematic storage as well as a tool for teaching/learning, and a tool for recalling knowledge, but it is also a neat way of pointing to knowledge expansion. It identifies voids, potential increments (or developments) in theory, and potential applications involving the existing theory. The basic motivations and uses for taxonomy may be listed as follows (Eksioglu et al. 2009):

- It defines or delimits the boundaries of a subject domain, and that is, in itself, useful information.
- It vividly, efficiently and effectively displays all of that domain's attributes/dimensions.
- It vividly, efficiently and effectively displays that any one of the possible combinations of these attributes/dimensions defines or delimits the boundaries of a subject sub-domain.
- It allows one to have a panoramic view of the entire "forest" while examining and classifying a given "tree."
- It allows one to unify disjointed and disparate subfields or sub-disciplines into a meaningful whole.
- It allows one to organize one's knowledge about the domain, and this has major implications for teaching, learning, storing, and recalling information.
- It allows one to identify voids and well explored territories in the extant literature base, which is very important for researchers, funding agencies, and other decision makers.

What is presented here is open for incremental evolution, as is the case in one of the greatest and best-known taxonomies of all time: The Periodic Table of Elements. The classification developed in this study is open to expansion when the scope of OR studies in healthcare management is enhanced, since any taxonomy is delimited with the boundaries of the universe it classifies.

OR studies in healthcare management have already generated a large enough literature to allow it to be considered as a separate and distinct field of knowledge. The increasing interest in the OR studies in healthcare management makes a systematic elaboration of this field more crucial to helping current researchers as well as attracting potential newcomers to the field.

Defining a taxonomy for OR Studies in Healthcare Management may seem to be overly detailed in terms of branching levels, as a result of trying to cover all literature in every subarea of healthcare management research. Although this detailed branching results in a taxonomy that is hard to work with, it increases its descriptive powers. Furthermore, it gives researchers the ability to aggregate sub-classifications and/or pruning outer branches easily. The taxonomy proceeds in a way illustrated by Reisman (1992), which can be seen in Fig. 1.2.

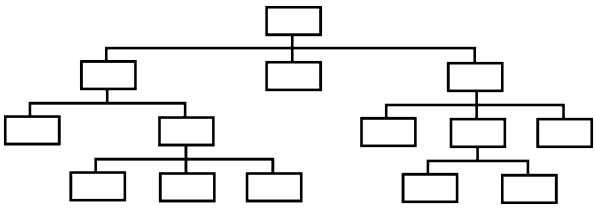

Fig. 1.2 Attribute vector description based taxonomy

1.4 A Taxonomy for OR Studies in Healthcare Management

In this section, the taxonomy for OR studies in Healthcare Management (HCM) is presented, and the main features that were considered while building it are introduced. We provide definitions as well as justifications for those main features and provide identification codes for some terms within the context of the taxonomy.

The full taxonomy is illustrated in Fig. 1.3. In the proposed taxonomy, each contribution can be given an identification code based on domains grouped in five classes:

Class 1: Study Specifications. This class shows how the study is specified. This is subdivided into three domains. The first domain describes the type of study; the second describes the source of the data used; and the third describes the type of problem treatment.

Class 2: Subject. This class shows what is analyzed. Each research paper analyzes one or more subjects. There are seven main subjects and the rest is grouped as "other".

Class 3: Methodology. This class shows the methodology used in the research. Each research paper consists of one or more methods. There are eleven main methods; and the rest is grouped as "other".

Class 4: Problem Specifications. This class shows who and what the problem is analyzed for. This is subdivided into three domains. The first domain describes the people affected by the problem; the second describes the area that the problem occurred in; and the third describes the affected facility by the problem.

Class 5: Location Specifications. This class shows where the research was carried out. The model constructed or the problem analyzed can be applied to large, medium or small scale; or it can be non-location-specific.

1.5 Results of the Taxonomy with Selected Articles

In this section, by using a group of articles which represent rather different approaches and which address different issues of OR studies in HCM, the taxonomy of Fig. 1.3 is tested for its robustness and its ability to discriminate in a rigorous manner.

1. Study Specifications
   1.1. Type of Study
      1.1.1. Model Construction using an Existing Method
      1.1.2. Model Construction using a Modified Method or Integration of Methods
      1.1.3. Method Comparison
   1.2. Data Used
      1.2.1. Real Data
      1.2.2. Both Real and Synthetic Data
   1.3. Problem Treatment
      1.3.1. Situation Analysis
      1.3.2. Decision Making (Problem Solving)
2. Subject
   2.1. Planning and Design
   2.2. Performance Measurement
   2.3. Capacity Management
   2.4. Scheduling and Assignment
   2.5. Resource/Budget Allocation
   2.6. Patient Flow and Waitlist Management
   2.7. Location
   2.8. Other
3. Methodology
   3.1. Linear/Integer Programming
   3.2. Multi Objective Programming
   3.3. Simulation
   3.4. Data Envelopment Analysis
   3.5. Queuing Theory
   3.6. System Dynamics
   3.7. Stochastic Methods
   3.8. Multi Attribute Decision Making
   3.9. Game Theory
   3.10. Bayesian Belief Network
   3.11. Artificial Neural Network
   3.12. Other
4. Problem Specifications
   4.1. Concerning People
      4.1.1. Management
      4.1.2. Doctor/Physician
      4.1.3. Nurse or Non-Medical Staff
      4.1.4. Patients
   4.2. Concerning Area
      4.2.1. Hospital/Clinic
      4.2.2. Non-hospital Organizations
      4.2.3. Public Health
   4.3. Concerning Facility
      4.3.1. Entire Clinic/Hospital
      4.3.2. Emergency Room
      4.3.3. Operating Room
      4.3.4. Ambulance
      4.3.5. Nursing Home
      4.3.6. Hospital Room
      4.3.7. Other
5. Location Specifications
   5.1. Large Scale
      5.1.1. Worldwide
      5.1.2 Continent Based
   5.2. Medium Scale
      5.2.1. Country Based
      5.2.2. State Based
   5.3. Small Scale
      5.3.1. City/Town Based
      5.3.2. Specific Location Based
   5.4. No Location Specific

Fig. 1.3 A taxonomy of OR studies in healthcare management

One hundred forty-two articles were investigated in detail to see the general idea of the researchers that contributed to the OR-in- HCM literature. In the first class, there are three domains; type of study, data used, and problem treatment. In the type of study domain, the most observed attribute is “model construction using a modified method” or “integration of methods” (1.1.2), followed by “model construction using an existing method” (1.1.1). “Comparison of methods” (1.1.3) is slightly less frequent than these two, since there are not enough studies to make a clear comparison. In data used, “usage of both real data and synthetic data” (1.2.2) is more frequent than “just using real data” (1.2.1), which can be explained by the difficulty of collecting real data as well as the highly popular usage of simulation that easily creates loads of synthetic data. For treating the problem, it is mostly “decision making” (1.3.2) rather than “situation analysis” (1.3.1). So, speaking for the study specifications, the papers mostly consisted of “decision making” with the “usage of both real and synthetic data” by “constructing a model using a modified method”.

The second and third classes are the classes that give researchers direction. It is better to interpret these two classes by looking at them together. Thus, the researcher may be able to pick the method to use for the subject he/she works on. However, first, one would need to check where previous studies have focused on. In the second class, "patient flow" and "waiting list management" (2.6) is the most researched subject. This is followed by "scheduling and assignment" (2.4), and "performance measurement" (2.2). "Resource/budget allocation" (2.5), "planning and design" (2.1), and "capacity management" (2.3) are more generalized subjects, for which researchers need to consider more factors, which means they are harder to model. Thus, they are not as attractive as the first three subjects. "Location" (2.7), and "other" (2.8) subjects have found fewer study areas than the rest, but these studies have been done mostly in recent years, which can be considered new research areas introduced to the discipline.

In class three, where methods are compared, "simulation" (3.3) is the most common method used to model in healthcare management, both alone or integrated with other methods. Simulation is mostly used to model "planning and design" (2.1), "scheduling and assignment" (2.4) and "patient flow and waiting list management" (2.6). Mathematical programming models such as "linear/integer programming" (3.1) and "multi-objective programming" (3.2) are also frequently used in order to model "scheduling and assignment" (2.4), "resource/budget allocation", (2.5) and "location" (2.7). Following common methods are "data envelopment analysis" (3.4), which is used mostly for "performance measurement" (2.2) and "stochastic methods" (3.7), mostly to model "capacity management" (2.3) problems. "Multi attribute decision making" (3.8), "game theory" (3.9) and "artificial neural networks" (3.11) are the least used methods as they have been introduced to healthcare operations management area in recent years.

Fourth class is where the problem details are explained. It includes three domains. In the first domain, concerning people, the most affected and investigated group in the papers is "management "(4.1.1). It is followed by "patients" (4.1.4), affected mostly in modeling "patient flow and waiting list management" (2.6) problems. "Doctor/physician" (4.1.2) and "nurse or non-medical staff" (4.1.3) groups are included generally in "scheduling and assignment" (2.4) problems. The second domain seeks whether the problem occurred inside or outside the hospital. Most of them are "hospital/clinic" (4.2.1) problems; the rest is "non-hospital" organizations (4.2.2) or "public health" (4.2.3). Concerning facility is the third domain in this class. Most of the studies include "the entire facility" (4.3.1). "Emergency room" (4.3.2) and "operating room" (4.3.3) are also important research areas for operations research methodology, especially for "linear/integer programming" (3.1) and "queuing theory" (3.5). "Ambulances" (4.3.4) and "nursing homes" (4.3.5) are the facilities that have been gaining importance in recent years.

The last class is the location where the research in a given paper is carried out. This resulted in "specific location based" (5.3.2) to have the highest frequency. "Country based" (5.2.1) is the second one, because researchers do research under the regulations of specific countries. "Worldwide" (5.1.1) and "continent based" (5.1.2) are the lowest location types that appear in these papers, since it is hard to construct a model that can be applied to a very large scale in a world with so many varieties.

1.6 Conclusions and Further Suggestions

Selection of papers for any taxonomy is a subjective work. The taxonomy described in this chapter tries to represent a variety of studies with different journals, different authors from different countries, differing paths to theory extension, differing application sectors and differing research strategies.

Being a new Management Science sub-discipline, the OR literature in healthcare management is growing exponentially like the other new sub-disciplines. This literature is recording advancements in theory and in solution methodology while at the same time expanding its domain of applications. When the previous bibliographic studies are compared with this taxonomy, it can be seen that new research areas are added to the discipline; new methods are used to model problems, and new approaches are applied to improve outputs. In spite of all these developments, research subjects are still divided as Fries (1976) stated; and simulation is still the most commonly utilized method to model problems as Jun et al. (1999) mentioned.

This taxonomy is formed with the motivation to determine application areas and specifications of OR studies in healthcare management as guidelines for future avenues of research. For future research, the most focused areas can be determined and the deficiencies in those areas can be satisfied with different approaches. For example, performance measurement problems are usually modeled with data envelopment analysis and from the management point of view. Therefore, using different methods or looking from a different angle, management can help eliminate the drawbacks of the previous studies. Or the least focused areas can be chosen to work on, such as performance measurement of the emergency room, which has not yet been studied. Also, with the addition of new areas as a result of the growing literature, improvements on this taxonomy can be made.

Appendix

The articles selected for the taxonomy can be seen in the following tables, with their classifications. The domains or attributes corresponding to endnodes are marked with ‘X’. Shaded columns represent domains or classes which branch, so that shading suggests why these columns are not marked. This representation scheme enables us to assign more designations in a confined space.

Table A.1 Summary of illustrative classifications of the first 70 articles (attributes 1–3.12)

	Article	1.	1.1.	1.1.1.	1.1.2.	1.1.3.	1.2.	1.2.1.	1.2.2.	1.3.	1.3.1.	1.3.2.	2.	2.1.	2.2.	2.3.	2.4.	2.5.	2.6.	2.7.	2.8.	3.	3.1.	3.2.	3.3.	3.4.	3.5.	3.6.	3.7.	3.8.	3.9.	3.10.	3.11.	3.12.
1	2000. Beaulieu et al.			x				x			x						x							x									
2	2000. Flessa			x				x			x							x						x									
3	2000. Rossetti et al.			x					x			x			x										x								
4	2000. Siddharthan et al.			x				x			x				x											x							
5	2001. Björkgren et al.			x				x			x				x			x								x							
6	2001. Congdon				x				x			x							x						x						x		
7	2001. Jacobs					x			x		x				x											x		x					
8	2001. Marshall et al.				x			x			x					x		x													x		
9	2001. Ratcliffe et al.			x					x		x							x	x						x								
10	2001. Swisher et al.				x				x		x			x											x								
11	2001. Vasilakis and El-Darzi			x					x		x					x			x						x								
12	2002. Chan et al.				x				x			x					x						x		x								
13	2002. Everett			x					x		x						x		x						x								
14	2002. Hofmarcher et al.			x				x			x				x											x							
15	2002. Kommer			x					x			x							x						x								
16	2002. Marshall et al.				x			x			x					x			x													x	
17	2002. Rauner			x					x		x			x				x							x								
18	2002. Rohleder and Klassen			x					x		x						x								x								
19	2002. Swisher and Jacobson			x					x			x		x	x										x								
20	2002. Verter and Lapierre			x				x				x								x			x										
21	2003. Chu et al.			x				x				x			x											x							
22	2003. De Angelis et al.				x				x		x			x									x		x								
23	2003. Flessa			x				x				x						x					x										
24	2003. Lane et al.			x				x			x								x								x						
25	2003. Mikkola et al.			x				x			x									x					x								
26	2003. Mullen					x						x					x		x														x
27	2003. Rauner and Bajmoczy				x			x				x			x								x										x
28	2003. Sendi and Al				x				x			x						x					x										
29	2004. Acid et al.					x			x		x				x																	x	
30	2004. Akkerman and Knip				x			x			x					x			x						x			x					
31	2004. Chang et al.			x				x				x			x											x							
32	2004. Grunow et al.			x				x				x					x	x					x										
33	2004. Gupta and Li			x					x			x								x			x										
34	2004. Hougaard et al.				x			x			x							x								x							
35	2004. Masterson et al.			x				x						x		x									x								
36	2004. Sibbritt and Gibberd				x				x		x										x												x
37	2004. Stummer et al.			x				x				x		x					x				x										
38	2005. Bard and Purnomo			x				x				x					x						x										
39	2005. Beynon and Kitchener				x				x			x			x															x			
40	2005. Harrison et al.				x				x		x					x									x				x				
41	2005. Koizumi et al.				x				x			x							x								x						
42	2005. Kontodimopoulos et al.			x				x			x				x											x							
43	2005. Marshall et al.					x						x				x			x						x				x				
44	2006. Aaby et. al				x				x			x		x		x									x		x						
45	2006. Akcali et al.			x					x			x				x							x										
46	2006. Bard and Purnomo				x				x			x					x						x										
47	2006. Cayirli et al.			x				x				x					x								x								
48	2006. Cochran and Bharti				x			x				x				x			x						x		x						
49	2006. Mitropoulos et al.			x				x				x								x				x									
50	2007. Aktas et al.			x				x			x							x													x		
51	2007. Anderson et al.			x					x		x										x				x								
52	2007. Cipriano et al.				x			x			x							x	x						x		x						
53	2007. de Bruin et al.			x				x			x					x			x								x						
54	2007. Denton et al.				x			x				x					x						x						x				
55	2007. Earnshaw et al.			x					x			x						x					x										
56	2007. Kopach et al.			x					x		x						x								x								
57	2007. Matta and Patterson			x				x			x				x										x								
58	2007. Molema et al.			x					x			x		x			x								x								
59	2007. Rohleder et al.			x					x			x		x													x						
60	2007. Santibáñez et al.			x				x				x					x		x				x										
61	2007. Testi et al.				x				x			x					x								x								
62	2007. VanBerkel and Blake			x					x		x					x			x						x								
63	2007. Vos et al.			x					x		x			x											x								
64	2008. Barros et al.				x			x			x				x								x										x
65	2008. Brasted			x				x				x							x						x								
66	2008. Clement et al.			x					x		x				x											x							
67	2008. Desai et al.			x					x			x		x													x						
68	2008. Dexter et al.				x				x		x										x					x							
69	2008. Goodson and Jang			x					x		x				x																	x	
70	2008. Ingolfsson et al.			x					x			x							x				x										

Table A.2 Summary of illustrative classifications of the first 70 articles (attributes 4–5.4)

	Article	4.	4.1.	4.1.1.	4.1.2.	4.1.3.	4.1.4.	4.2.	4.2.1.	4.2.2.	4.2.3.	4.3.	4.3.1.	4.3.2.	4.3.3.	4.3.4.	4.3.5.	4.3.6.	4.3.7.	5.	5.1.	5.1.1.	5.1.2.	5.2.	5.2.1.	5.2.2.	5.3.	5.3.1.	5.3.2.	5.4.
1	2000. Beaulieu et al.				x				x					x											x				
2	2000. Flessa			x							x								x			x							
3	2000. Rossetti et al.			x					x				x																x
4	2000. Siddharthan et al.			x						x									x					x					
5	2001. Björkgren et al.			x						x			x											x					
6	2001. Congdon						x				x			x										x					
7	2001. Jacobs			x					x				x											x					
8	2001. Marshall et al.			x					x										x									x	
9	2001. Ratcliffe et al.						x			x									x									x	
10	2001. Swisher et al.			x					x				x															x	
11	2001. Vasilakis and El-Darzi						x		x									x										x	
12	2002. Chan et al.					x			x										x									x	
13	2002. Everett						x		x						x														x
14	2002. Hofmarcher et al.			x					x				x														x		
15	2002. Kommer						x				x								x					x					
16	2002. Marshall et al.			x			x		x									x										x	
17	2002. Rauner			x							x																x		
18	2002. Rohleder and Klassen						x		x										x										x
19	2002. Swisher and Jacobson			x					x				x															x	
20	2002. Verter and Lapierre			x					x				x															x	
21	2003. Chu et al.				x				x															x					
22	2003. De Angelis et al.			x					x				x											x					
23	2003. Flessa			x							x														x				
24	2003. Lane et al.						x		x					x														x	
25	2003. Mikkola et al.										x													x					
26	2003. Mullen																												x
27	2003. Rauner and Bajmoczy			x						x									x					x					
28	2003. Sendi and Al			x																									x
29	2004. Acid et al.			x					x					x														x	
30	2004. Akkerman and Knip						x		x						x									x					
31	2004. Chang et al.			x					x				x											x					
32	2004. Grunow et al.				x					x									x										x
33	2004. Gupta and Li						x			x														x					
34	2004. Hougaard et al.						x				x						x		x					x					
35	2004. Masterson et al.			x						x									x					x					
36	2004. Sibbritt and Gibberd			x					x										x									x	
37	2004. Stummer et al.			x					x				x														x		
38	2005. Bard and Purnomo					x			x				x															x	
39	2005. Beynon and Kitchener									x														x					
40	2005. Harrison et al.						x		x									x										x	
41	2005. Koizumi et al.						x		x																x				
42	2005. Kontodimopoulos et al.			x					x				x											x					
43	2005. Marshall et al.						x		x									x											x
44	2006. Aaby et. al			x							x			x													x		
45	2006. Akcali et al.						x		x									x											x
46	2006. Bard and Purnomo					x			x																				x
47	2006. Cayirli et al.				x		x		x										x									x	
48	2006. Cochran and Bharti						x		x									x						x					
49	2006. Mitropoulos et al.			x					x																x				
50	2007. Aktas et al.			x					x										x									x	
51	2007. Anderson et al.			x							x											x							
52	2007. Cipriano et al.						x		x						x										x				
53	2007. de Bruin et al.						x		x					x														x	
54	2007. Denton et al.								x						x													x	
55	2007. Earnshaw et al.			x							x														x				
56	2007. Kopach et al.				x		x		x																				x
57	2007. Matta and Patterson			x					x				x															x	
58	2007. Molema et al.			x	x				x						x				x									x	
59	2007. Rohleder et al.			x						x																		x	
60	2007. Santibáñez et al.				x		x		x						x													x	
61	2007. Testi et al.								x						x													x	
62	2007. VanBerkel and Blake						x		x						x													x	
63	2007. Vos et al.			x					x				x															x	
64	2008. Barros et al.			x					x				x											x					
65	2008. Brasted						x		x										x									x	
66	2008. Clement et al.			x					x				x											x					
67	2008. Desai et al.						x				x						x										x		
68	2008. Dexter et al.				x				x						x														x
69	2008. Goodson and Jang									x							x							x					
70	2008. Ingolfsson et al.									x						x											x		

Table A.3 Summary of illustrative classifications of the last 72 articles (attributes 1–3.12)

	Article	1.	1.1.	1.1.1.	1.1.2.	1.1.3.	1.2.	1.2.1.	1.2.2.	1.3.	1.3.1.	1.3.2.	2.	2.1.	2.2.	2.3.	2.4.	2.5.	2.6.	2.7.	2.8.	3.	3.1.	3.2.	3.3.	3.4.	3.5.	3.6.	3.7.	3.8.	3.9.	3.10.	3.11.	3.12.
71	2008. Jiang and Giachetti					x		x				x							x						x		x						
72	2008. Lim and Kirikoshi				x			x			x									x												x	
73	2008. Little and Coughlan			x				x				x						x						x									
74	2008. Mayhew and Smith			x					x		x								x								x						
75	2008. Mazur and Chen				x			x				x		x																			x
76	2008. Puenpatom and Rosenman			x				x			x				x											x							
77	2008. Sobolev et al.				x				x		x								x						x								
78	2008. Utley et al.				x			x				x				x												x					
79	2008. van de Klundert et al.				x				x			x		x									x										x
80	2009. Adan et al.				x			x				x					x	x					x					x					
81	2009. Augusto and Xie				x			x				x		x					x				x		x								
82	2009. Brunner et al.			x				x				x					x						x										
83	2009. Chand et al.				x			x				x							x						x								x
84	2009. De Grano et al.				x			x				x					x							x									
85	2009. Hare et al.				x				x		x					x												x					
86	2009. Lavieri and Puterman			x				x				x		x			x						x										
87	2009. Marjamaa et al.			x					x			x		x			x								x								
88	2009. Restrepo et al.				x				x			x						x		x			x										
89	2009. Santibáñez et al.			x				x				x						x	x						x								
90	2009. Sharma			x				x			x							x	x										x				
91	2009. Sundaramoorthi et al.				x				x		x						x								x				x				
92	2009. Testi and Tànfani			x					x			x		x									x										
93	2009. Tiwari and Heese			x					x			x		x																x			
94	2010. Blank and Valdmanis				x			x			x				x											x							
95	2010. Garg et al.				x				x		x								x									x					
96	2010. Joustra et al.				x			x				x							x				x		x								
97	2010. Kristensen et al.			x				x			x				x											x							
98	2010. Linna et al.			x				x			x				x											x							
99	2010. Mukherjee et al.			x				x			x				x											x							
100	2010. Persson and Persson			x				x				x						x		x					x								
101	2010. Rönnberg and Larsson			x				x				x						x					x										
102	2010. Sundaramoorthi et al.					x		x				x						x							x								
103	2010. Varela et al.			x				x			x				x											x							
104	2010. Yaesoubi and Roberts				x				x		x							x			x									x			
105	2010. Zhang et al.				x			x				x		x						x				x									
106	2010. Zonderland et al.				x				x			x						x									x		x				
107	2011a. Büyüközkan et al.				x				x			x			x														x				
108	2011. Carpenter et al.				x			x				x					x	x					x										
109	2011. Conforti et al.			x				x				x					x						x										
110	2011. Dursun et al				x			x				x		x															x				
111	2011. Garavaglia et al.				x			x			x				x											x							
112	2011. Kreng and Young			x				x				x						x															x
113	2011. Lasry et al.					x			x		x							x					x										x
114	2011. Rohleder et al.				x			x				x							x						x								
115	2012. Büyüközkan and Çifçi				x				x			x			x															x				
116	2012. Defechereux at al			x					x		x			x															x				
117	2012. Goetghebeur et al			x				x			x										x								x				
118	2012. Grigoroudis et al.			x					x			x			x															x				
119	2012. Kuo et al				x			x				x		x																			x
120	2012. Thokala and Duenas					x			x			x									x									x				
121	2012. Zhang and Huang				x				x		x			x																	x			x
122	2013. Abo-Hamad and Arisha				x			x			x				x				x						x								
123	2013. Chen et al				x				x			x									x									x				x
124	2013. Hulshof et al			x					x			x							x				x											
125	2013. Marsh et al			x					x			x		x																x				
126	2013. Özkan					x			x			x									x									x				
127	2013. Toro-Díaz et al				x				x			x						x		x													x	
128	2013. Zeng et al				x			x				x		x																x				
129	2014. Akdag et al					x			x		x				x															x				
130	2014. Amaral and Costa			x					x			x							x											x				
131	2014. Choi and Wilhelm				x				x			x					x							x										
132	2014. Lashgari et al					x			x			x				x														x				
133	2014. Mitton et al				x						x							x																x
134	2014. Robinson et al			x					x		x				x										x									
135	2014. Sun et al				x			x				x						x						x										
136	2015. Khaki et al				x				x			x								x										x				x
137	2015. Hernandez et al				x				x			x					x							x	x									
138	2015. Kanuganti et al				x				x		x									x										x				x
139	2015. Girginer et al				x				x		x				x											x								x
140	2015. Rashwan et al			x				x			x					x											x							
141	2015. McCormack and Coates				x			x				x						x							x								x	
142	2015. Xu et al					x			x		x				x												x		x					x

Table A.4 Summary of illustrative classifications of the last 72 articles (attributes 4–5.4)

	Article	4.	4.1.	4.1.1.	4.1.2.	4.1.3.	4.1.4.	4.2.	4.2.1.	4.2.2.	4.2.3.	4.3.	4.3.1.	4.3.2.	4.3.3.	4.3.4.	4.3.5.	4.3.6.	4.3.7.	5.	5.1.	5.1.1.	5.1.2.	5.2.	5.2.1.	5.2.2.	5.3.	5.3.1.	5.3.2.	5.4.
71	2008. Jiang and Giachetti						x			x									x										x
72	2008. Lim and Kirikoshi				x					x									x					x					
73	2008. Little and Coughlan			x					x										x									x	
74	2008. Mayhew and Smith			x							x			x										x					
75	2008. Mazur and Chen			x					x										x									x	
76	2008. Puenpatom and Rosenman			x					x		x		x											x					
77	2008. Sobolev et al.						x		x						x									x					
78	2008. Utley et al.			x						x								x						x					
79	2008. van de Klundert et al.			x					x						x				x										x
80	2009. Adan et al.			x					x						x													x	
81	2009. Augusto and Xie					x				x									x									x	
82	2009. Brunner et al.				x				x																			x	
83	2009. Chand et al.						x		x																			x	
84	2009. De Grano et al.								x					x														x	
85	2009. Hare et al.						x			x							x								x				
86	2009. Lavieri and Puterman					x					x														x				
87	2009. Marjamaa et al.			x	x	x			x						x													x	
88	2009. Restrepo et al.			x						x						x											x		
89	2009. Santibáñez et al.			x			x			x			x															x	
90	2009. Sharma						x		x						x									x					
91	2009. Sundaramoorthi et al.					x	x		x				x															x	
92	2009. Testi and Tànfani			x					x						x													x	
93	2009. Tiwari and Heese			x					x				x																x
94	2010. Blank and Valdmanis			x					x				x											x					
95	2010. Garg et al.						x																						x
96	2010. Joustra et al.						x		x										x									x	
97	2010. Kristensen et al.			x					x															x					
98	2010. Linna et al.			x					x		x		x										x						
99	2010. Mukherjee et al.			x							x													x					
100	2010. Persson and Persson				x		x		x						x													x	
101	2010. Rönnberg and Larsson					x			x										x									x	
102	2010. Sundaramoorthi et al.					x	x		x									x										x	
103	2010. Varela et al.			x							x														x				
104	2010. Yaesoubi and Roberts						x		x	x														x					
105	2010. Zhang et al.			x						x			x												x				
106	2010. Zonderland et al.			x			x		x						x													x	
107	2011a. Büyüközkan et al.			x					x				x														x		
108	2011. Carpenter et al.					x			x										x									x	
109	2011. Conforti et al.						x		x				x															x	
110	2011. Dursun et al			x						x									x								x		
111	2011. Garavaglia et al.			x						x							x							x					
112	2011. Kreng and Young			x							x													x					
113	2011. Lasry et al.			x						x														x					
114	2011. Rohleder et al.						x		x				x															x	
115	2012. Büyüközkan and Çifçi			x					x										x										x
116	2012. Defechereux at al			x							x													x					
117	2012. Goetghebeur et al			x						x														x					
118	2012. Grigoroudis et al.			x					x		x		x															x	
119	2012. Kuo et al						x				x						x							x					
120	2012. Thokala and Duenas			x							x													x					
121	2012. Zhang and Huang			x							x																x		
122	2013. Abo-Hamad and Arisha			x					x					x															x
123	2013. Chen et al				x				x										x									x	
124	2013. Hulshof et al			x					x				x																x
125	2013. Marsh et al			x							x													x					
126	2013. Özkan			x						x									x					x					
127	2013. Toro-Díaz et al			x					x					x															x
128	2013. Zeng et al			x							x													x					
129	2014. Akdag et al								x				x											x					
130	2014. Amaral and Costa			x					x					x														x	
131	2014. Choi and Wilhelm			x					x						x														x
132	2014. Lashgari et al			x						x			x															x	
133	2014. Mitton et al			x							x										x								
134	2014. Robinson et al			x					x				x															x	
135	2014. Sun et al						x		x					x											x				
136	2015. Khaki et al										x			x											x				
137	2015. Hernandez et al				x	x					x			x													x		
138	2015. Kanuganti et al						x		x				x												x				
139	2015. Girginer et al			x					x						x												x		
140	2015. Rashwan et al			x							x			x										x					
141	2015. McCormack and Coates						x									x											x		
142	2015. Xu et al			x					x				x														x		

References

Aaby K, Herrmann JW, Jordan CS, Treadwell M, Wood K (2006) Montgomery county's public health service uses operations research to plan emergency mass dispensing and vaccination clinics. Interfaces 36:569–579

Abo-Hamad W, Arisha A (2013) Simulation-based framework to improve patient experience in an emergency department. Eur J Oper Res 224:154–166

Acid S, de Campos LM, Fernandez-Luna JM, Rodriguez S, Rodriguez JM, Salcedo JL (2004) A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service. Artif Intell Med 30:215–232

Adan I, Bekkers J, Dellaert N, Vissers J, Yu X (2009) Patient mix optimisation and stochastic resource requirements: a case study in cardiothoracic surgery planning. Health Care Manag Sci 12:129–141

Akcali E, Cote MJ, Lin C (2006) A network flow approach to optimizing hospital bed capacity decisions. Health Care Manag Sci 9:391–404

Akdag H, Kalaycı T, Karagöz S, Zülfikar H, Giz D (2014) The evaluation of hospital service quality by fuzzy MCDM. Appl Soft Comput 23:239–248

Akkerman R, Knip M (2004) Reallocation of beds to reduce waiting time for cardiac surgery. Health Care Manag Sci 7:119–126

Aktas E, Ulengin F, Sahin SO (2007) A decision support system to improve the efficiency of resource allocation in healthcare management. Socio Econ Plan Sci 41:130–146

Amaral TM, Costa APC (2014) Improving decision-making and management of hospital resources: an application of the PROMETHEE II method in an emergency department. Oper Res Health Care 3:1–6

Anderson J, Chaturvedi A, Cibulskis M (2007) Simulation tools for developing policies for complex systems: modeling the health and safety of refugee communities. Health Care Manag Sci 10:331–339

Augusto V, Xie X (2009) Redesigning pharmacy delivery processes of a health care complex. Health Care Manag Sci 12:166–178

Bard JF, Purnomo HW (2006) Incremental changes in the workforce to accommodate changes in demand. Health Care Manag Sci 9:71–85

Barros CP, de Menezes AG, Peypoch N, Solonandrasana B, Vieira JC (2008) An analysis of hospital efficiency and productivity growth using the Luenberger indicator. Health Care Manag Sci 11:373–381

Beaulieu H, Ferland JA, Gendron B, Michelon P (2000) A mathematical programming approach for scheduling physicians in emergency room. Health Care Manag Sci 3:193–200

Björkgren MA, Häkkinen U, Linna M (2001) Measuring efficiency of long-term care units in Finland. Health Care Manag Sci 4:193–200

Blank JLT, Valdmanis VG (2010) Environmental factors and productivity on Dutch hospitals: a semi-parametric approach. Health Care Manag Sci 13:27–34

Brasted C (2008) Ultrasound waiting lists: rational queue or extended capacity? Health Care Manag Sci 11:196–207

Brunner JO, Bard JF, Kolisch R (2009) Flexible shift scheduling of physicians. Health Care Manag Sci 12:285–305

Buyukozkan G, Çifçi G (2012) A combined fuzzy AHP and fuzzy TOPSIS based strategic analysis of electronic service quality in healthcare industry. Expert Syst Appl 39(3):2341–2354

Büyüközkan G, Çiftçi G, Güleryüz S (2011) Strategic analysis of healthcare service quality using fuzzy AHP methodology. Expert Syst Appl 38:9407–9424

Carpenter AP, Leemis LM, Papir AS, Phillips DJ, Phillips GS (2011) Managing magnetic resonance imaging machines: support tools for scheduling and planning. Health Care Manag Sci 14:158–173

Carter M (2002) Diagnosis: mismanagement of resources. OR/MS Today 29:26–33

Cayirli T, Veral E (2003) Outpatient scheduling in health care: a review of literature. Prod Oper Manag 12:519–549

Cayirli T, Veral E, Rosen H (2006) Designing appointment scheduling systems for ambulatory care services. Health Care Manag Sci 9:47–58

Chan SY, Ohlmann J, Dunbar S, Dunbar C, Ryan S, Savory P (2002) Operations research methods applied to workflow in a medical records department. Health Care Manag Sci 5:191–199

Chand S, Moskowitz H, Norris JB, Shade S, Willis DR (2009) Improving patient flow at an outpatient clinic: study of sources of variability and improvement factors. Health Care Manag Sci 12:325–340

Chang H, Cheng MA, Das S (2004) Hospital ownership and operating efficiency: evidence from Taiwan. Eur J Oper Res 159:513–527

Chen T-Y, Chang C-H, Lu JF (2013) The extended QUALIFLEX method for multiple criteria decision analysis based on interval type-2 fuzzy sets and applications to medical decision making. Eur J Oper Res 226:615–625

Choi S, Wilhelm WE (2014) An approach to optimize block surgical schedules. Eur J Oper Res 235:138–148

Chu HL, Liu SZ, Romeis JC, Yaung CL (2003) The initial effects of physician compensation programs in Taiwan hospitals: implications for staff model HMOs. Health Care Manag Sci 6:17–26

Cipriano LE, Chesworth BM, Anderson CK, Zaric GS (2007) Predicting joint replacement waiting times. Health Care Manag Sci 10:195–215

Clement JP, Valdmanis VG, Bazzoli GJ, Zhao M, Chukmaitov A (2008) Is more better? An analysis of hospital outcomes and efficiency with a DEA model of output congestion. Health Care Manag Sci 11:67–77

Cochran JK, Bharti A (2006) Stochastic bed balancing of an obstetrics hospital. Health Care Manag Sci 9:31–45

Congdon P (2001) The development of gravity models for hospital patient flows under system change: a Bayesian modelling approach. Health Care Manag Sci 4:289–304

Conforti D, Guerriero F, Guido R, Cerinic MM, Conforti ML (2011) An optimal decision making model for supporting week hospital management. Health Care Manag Sci 14:74–88

Corner JL, Kirkwood CW (1991) Decision analysis applications in the operations research literature. Oper Res 39:206–219

De Angelis V, Felici G, Impelluso P (2003) Integrating simulation and optimisation in health care centre management. Eur J Oper Res 150:101–114

de Bruin AM, van Rossum AC, Visser MC, Koole GM (2007) Modeling the emergency cardiac in-patient flow: application of queuing theory. Health Care Manag Sci 10:125–137

De Grano ML, Medeiros DJ, Eitel D (2009) Accommodating individual preferences in nurse scheduling via auctions and optimization. Health Care Manag Sci 12:228–242

Defechereux T, Paolucci F, Mirelman A, Youngkong S, Botten G, Hagen TP, Niessen LW. Health care priority setting in Norway a multicriteria decision analysis. BMC Health Serv Res. 2012;12(39), viewed 26 Feb 2016. https://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-12-39.

Denton B, Viapiano J, Vogl A (2007) Optimization of surgery sequencing and scheduling decisions under uncertainty. Health Care Manag Sci 10:13–24

Desai MS, Penn ML, Brailsford S, Chipulu M (2008) Modelling of Hampshire Adult Services—gearing up for future demands. Health Care Manag Sci 11:167–176

Dexter F, O'Neill L, Xin L, Ledolter J (2008) Sensitivity of super-efficient data envelopment analysis results to individual decision-making units: an example of surgical workload by specialty. Health Care Manag Sci 11:307–318

Dursun M, Karsak EE, Almula Karadayi M (2011) A fuzzy multi-criteria group decision making framework for evaluating health-care waste disposal alternatives. Expert Syst Appl 38(9):11453–11462

Earnshaw SR, Hicks K, Richter A, Honeycutt A (2007) A linear programming model for allocating HIV prevention funds with state agencies: a pilot study. Health Care Manag Sci 10:239–252

Eksioglu B, Vural V, Reisman A (2009) The vehicle routing problem: a taxonomic review. Comput Ind Eng 57:1472–1483
Everett JE (2002) A decision support simulation model for the management of an elective surgery waiting system. Health Care Manag Sci 5:89–95
Fries BE (1976) Bibliography of operations research. Oper Res 24:801–814
Fries BE (1981) Applications of operations research to health care delivery systems: a complete review of periodical literature. Springer-Verlag, Berlin
Flagle CD (1962) Operations research in the health services. The Johns Hopkins University and Hospital, Baltimore
Flessa S (2000) Where efficiency saves lives: a linear programme for optimal allocation of health care resources in developing countries. Health Care Manag Sci 3:249–267
Flessa S (2003) Priorities and allocation of health care resources in developing countries: a case-study from the Mtwara region, Tanzania. Eur J Oper Res 150:67–80
Garg L, McClean S, Meenan B, Millard P (2010) A non-homogeneous discrete time Markov model for admission scheduling and resource planning in a cost or capacity constrained healthcare system. Health Care Manag Sci 13:155–169
Garavaglia G, Lettieri E, Agasisti T, Lopez S (2011) Efficiency and quality of care in nursing homes: an Italian case study. Health Care Manag Sci 14:22–35
Gattoufi S, Oral M, Reisman A (2009) A taxonomy for data envelopment analysis. Socio Econ Plan Sci 38:141–158
Girginer N, Köse T, Uçkun N (2015) Efficiency analysis of surgical services by combined use of data envelopment analysis and gray relational analysis. J Med Syst 39:56
Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D (2012) Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal. Med Decis Making 32:376–388
Goodson J, Jang W (2008) Assessing nursing home care quality through Bayesian networks. Health Care Manag Sci 11:382–392
Grigoroudis E, Orfanoudaki E, Zopounidis E (2012) Strategic performance measurement in a healthcare organisation: a multiple criteria approach based on balanced scorecard. Omega 40:104–119
Grunow M, Günther HO, Yang G (2004) Development of a decision support model for scheduling clinical studies and assigning medical personnel. Health Care Manag Sci 7:305–317
Gupta A, Li L (2004) A modeling framework for optimal long-term care insurance purchase decisions in retirement planning. Health Care Manag Sci 7:105–117
Hare WL, Alimadad A, Dodd H, Ferguson R, Rutherford A (2009) A deterministic model of home and community care client counts in British Columbia. Health Care Manag Sci 12:80–98
Harrison GW, Shafer A, Mackay M (2005) Modelling variability in hospital bed occupancy. Health Care Manag Sci 8:325–334
Hernandez I, Ramirez-Marquez JE, Starr D, McKay R, Guthartz S, Motherwell M, Barcellona J (2015) Optimal staffing strategies for points of dispensing. Comput Ind Eng 83:172–183
Hillier FS, Lieberman GJ (2005) Introduction to operations research, 8th. (International) edn. McGraw-Hill, Boston
Hofmarcher MM, Paterson I, Riedel M (2002) Measuring hospital efficiency in Austria – a DEA approach. Health Care Manag Sci 5:7–14
Hougaard JL, Kronborg D, Overgard C (2004) Improvement potential in Danish elderly care. Health Care Manag Sci 7:225–235
Hulshof PJH, Kortbeek N, Boucherie RJ, Hans EW, Bakker PJM (2012) Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. Health Syst 1:129–175
Hulshof PJH, Boucherie RJ, Hans EW, Hurink JL (2013) Tactical resource allocation and elective patient admission planning in care processes. Health Care Manag Sci 16:152–166
Ingolfsson A, Budge S, Erkut E (2008) Optimal ambulance location with random delays and travel times. Health Care Manag Sci 11:262–274

Jacobs R (2001) Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis. Health Care Manag Sci 4:103–115
Jiang L, Giachetti RE (2008) A queueing network model to analyze the impact of parallelization of care on patient cycle time. Health Care Manag Sci 11:248–261
Joustra PE, de Wit J, Struben VMD, Overbeek BJH, Fockens P, Elkhuizen SG (2010) Reducing access times for an endoscopy department by an iterative combination of computer simulation and linear programming. Health Care Manag Sci 13:17–26
Jun JB, Jacobson SH, Swisher JR (1999) Application of discrete-event simulation in health care clinics: a survey. J Oper Res Soc 50:109–123
Kanuganti S, Sarkar AK, Singh AP, Arkatkar SS (2015) Quantification of accessibility to health facilities in rural areas. Case Stud Transp Policy 3:311–320
Khaki AM, Mojaradi B, Ghobadipour B, Maghsoudy S, Naghibi F (2015) Integration of GIS and analytical hierarchy process method for locating road emergency medical services station. Geosyst Eng 18(2):92–103
Kirby MW (2003) Operational research in war and peace: the British experience from the 1930s to 1970. Imperial College Press, London
Koizumo N, Kuno E, Smith TE (2005) Modeling patient flows using a queuing network with blocking. Health Care Manag Sci 8:49–60
Kommer GJ (2002) A waiting list model for residential care for the mentally disabled in the Netherlands. Health Care Manag Sci 5:285–290
Kontodimopoulos N, Nanos P, Niakas D (2006) Balancing efficiency of health services and equity of access in remote areas in Greece. Health Policy 76:49–57
Kopach R, DeLaurentis P, Lawley M, Muthuraman K, Ozsen L, Rardin R, Wan H, Intrevado P, Qu X, Willis D (2007) Effects of clinical characteristics on successful open access scheduling. Health Care Manag Sci 10:111–124
Kreng VB, Yang C-T (2011) The equality of resource allocation in health care under the National Health Insurance System in Taiwan. Health Policy 100(2):203–210
Kristensen T, Bogetoft P, Pedersen KM (2010) Potential gains from hospital mergers in Denmark. Health Care Manag Sci 13:334–345
Kuo R-J, Wu Y-H, Hsu T-S (2012) Integration of fuzzy set theory and TOPSIS into HFMEA to improve outpatient service for elderly patients in Taiwan. J Chin Med Assoc 75:341–348
Lane DC, Monefeldt C, Husemann E (2003) Client involvement in simulation model building: hints and insights from a case study in a London hospital. Health Care Manag Sci 6:105–116
Langabeer J (2007) Health care operations management: a quantitative approach to business and logistics, 1st edn. Jones and Bartlett Publishers, Sudbury
Lashgari S, Antuchevičienė J, Delavari A, Kheirkhah O (2014) Using QSPM and WASPAS methods for determining outsourcing strategies. J Bus Econ Manag 15(4):729–743
Lasry A, Sansom SL, Hicks KA, Uzunangelov V (2011) A model for allocating CDC's HIV prevention resources in the United States. Health Care Manag Sci 14:115–124
Lavieri MS, Puterman ML (2009) Optimizing nursing human resource planning in British Columbia. Health Care Manag Sci 12:119–128
Lim CW, Kirikoshi T (2008) Understanding the effects of pharmaceutical promotion: a neural network approach guided by genetic algorithm-partial least squares. Health Care Manag Sci 11:359–372
Linna M, Häkkinen U, Peltola M, Magnussen J, Anthun KJ, Kittelsen S, Roed A, Olsen K, Medin E, Rehnberg C (2010) Measuring cost efficiency in the Nordic Hospitals—a cross-sectional comparison of public hospitals in 2002. Health Care Manag Sci 13:346–357
Little J, Coughlan B (2008) Optimal inventory policy within hospital space constraints. Health Care Manag Sci 11:177–183
Lowery JC (1996) Introduction to simulation in health care. Proceedings of the 1996 Winter Simulation Conference, pp 78–84
Luss H, Rosenwein M (1997) Operations research applications: opportunities and accomplishments. Eur J Oper Res 97:220–244

Marjamaa RA, Torkki PM, Hirvensalo EJ, Kirvelä OA (2009) What is the best workflow for an operating room? A simulation study of five scenarios. Health Care Manag Sci 12:142–146
Marsh K, Dolan P, Kempster J, Lugon M (2013) Prioritizing investments in public health: a multi-criteria decision analysis. J Public Health 35(3):460–466
Marshall A, Vasilakis C, El-Darzi E (2005) Length of stay-based patient flow models: recent developments and future directions. Health Care Manag Sci 8:213–220
Marshall AH, McClean SI, Shapcott CM, Hastie IR, Millard PH (2001) Developing a Bayesian belief network for the management of geriatric hospital care. Health Care Manag Sci 4:25–30
Marshall AH, McClean SI, Shapcott CM, Millard PH (2002) Modelling patient duration of stay to facilitate resource management of geriatric hospitals. Health Care Manag Sci 5:313–319
Masterson BJ, Mihara TG, Miller G, Randolph SC, Forkner ME, Crouter AL (2004) Using models and data to support optimization of the military health system: a case study in an intensive care unit. Health Care Manag Sci 7:217–224
Matta ME, Patterson SS (2007) Evaluating multiple performance measures across several dimensions at a multi-facility outpatient center. Health Care Manag Sci 10:173–194
Mayhew L, Smith D (2008) Using queuing theory to analyse the government's 4-h completion time target in accident and emergency departments. Health Care Manag Sci 11:11–21
Mazur LM, Chen SJ (2008) Understanding and reducing the medication delivery waste via systems mapping and analysis. Health Care Manag Sci 11:55–65
McCormack R, Coates G (2015) A simulation model to enable the optimization of ambulance fleet allocation and base station location for increased patient survival. Eur J Oper Res 247:294–309
Mclaughlin DB, Hays JM (2008) Healthcare operations management. Health Administration Press, Chicago
Mikkola H, Sund R, Linna M, Häkkinen U (2003) Comparing the financial risk of bed-day and DRG based pricing types using parametric and simulation methods. Health Care Manag Sci 6:67–74
Mitropoulos P, Mitropoulos I, Giannikos I, Sissouras A (2006) A biobjective model for the locational planning of hospitals and health centers. Health Care Manag Sci 9:171–179
Mitton C, Dionne F, Donaldson C (2014) Managing healthcare budgets in times of austerity: the role of program budgeting and marginal analysis. Appl Health Econ Health Policy 12:95–102
Molema JJW, Groothuis S, Baars IJ, Kleinschiphorst M, Leers EGE, Hasman A, van Merode GG (2007) Healthcare system design and parttime working doctors. Health Care Manag Sci 10:365–371
Mukherjee K, Santerre RE, Zhang NJ (2010) Explaining the efficiency of local health departments in the U.S.: an exploratory analysis. Health Care Manag Sci 13:378–387
Mullen PM (2003) Prioritising waiting lists: how and why? Eur J Oper Res 150:32–45
Özkan A (2013) Evaluation of healthcare waste treatment/disposal alternatives by using multi-criteria decision-making techniques. Waste Manag Res 31(2):141–149
Persson MJ, Persson JA (2010) Analysing management policies for operating room planning using simulation. Health Care Manag Sci 13:182–191
Pierskalla WP, Brailer D (1994) Applications of operations research in health care delivery. In: Pollock S, Barnett A, Rothkopf M (eds) Beyond the profit motive: public sector applications and methodology, Handbooks in OR&MS, vol 6. Elsevier, North Holland, pp 469–505
Preater J (2002) Queues in health. Health Care Manag Sci 5:283
Puenpatom RA, Rosenman R (2008) Efficiency of Thai provincial public hospitals during the introduction of universal health coverage using capitation. Health Care Manag Sci 11:319–338
Rashwan W, Abo-Hamad W, Arisha A (2015) A system dynamics view of the acute bed blockage problem in the Irish healthcare system. Eur J Oper Res 247:276–293
Ratcliffe J, Young T, Buxton M, Eldabi T, Paul R, Burroughs A, Papatheodoridis G, Rolles K (2001) A simulation modelling approach to evaluating alternative policies for the management of the waiting list for liver transplantation. Health Care Manag Sci 4:117–124
Rauner MS (2002) Using simulation for AIDS policy modeling: benefits for HIV/AIDS prevention policy makers in Vienna. Health Care Manag Sci 5:121–134

Rauner MS, Bajmoczy N (2003) How many AEDs in which region? An economic decision model for the Austrian Red Cross. Eur J Oper Res 150:3–18
Reisman A (1992) Management science knowledge: it's creation generalization and consolidation. Quorum Books Publishing Company, Westport
Reisman A (1993) Operations management: time for meta research. Prospectives in operations management: essays in honor of Elwood S. Buffa. Springer, Boston, pp 471–482
Restrepo M, Henderson SG, Topaloglu H (2009) Erlang loss models for the static deployment of ambulances. Health Care Manag Sci 12:67–79
Robinson S, Worthington C, Burgess N, Radnor ZJ (2014) Facilitated modelling with discrete-event simulation: reality or myth? Eur J Oper Res 234:231–240
Rohleder TR, Bischak DP, Baskin LB (2007) Modeling patient service centers with simulation and system dynamics. Health Care Manag Sci 10:1–12
Rohleder TR, Lewkonia P, Bischak DP, Duffy P, Hendijani R (2011) Using simulation modeling to improve patient flow at an outpatient orthopedic clinic. Health Care Manag Sci 14:135–145
Rohleder TS, Klassen KJ (2002) Rolling horizon appointment scheduling: a simulation study. Health Care Manag Sci 5:201–209
Rossetti MD, Felder RA, Kumar A (2000) Simulation of robotic courier deliveries in hospital distribution services. Health Care Manag Sci 3:201–213
Rönnberg E, Larsson T (2010) Automating the self-scheduling process of nurses in Swedish healthcare: a pilot study. Health Care Manag Sci 13:35–53
Santibáñez P, Begen M, Atkins D (2007) Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a British Columbia health authority. Health Care Manag Sci 10:269–282
Santibáñez P, Chow VS, French J, Puterman ML, Tyldesley S (2009) Reducing patient wait times and improving resource utilization at British Columbia Cancer Agency's ambulatory care unit through simulation. Health Care Manag Sci 12:392–407
Sendi P, Al MJ (2003) Revisiting the decision rule of cost–effectiveness analysis under certainty and uncertainty. Soc Sci Med 57:969–974
Sharma C (2009) Inter-DRG resource dynamics in a prospective payment system: a stochastic kernel approach. Health Care Manag Sci 12:38–55
Shuman J, Speas RD Jr, Young JP (1975) Operations research in health care: a critical analysis. Johns Hopkins University Press, Baltimore
Siddharthan K, Ahern M, Rosenman R (2000) Data envelopment analysis to determine efficiencies of health maintenance organizations. Health Care Manag Sci 3:23–29
Sibbritt D, Gibberd R (2004) The effective use of a summary table and decision tree methodology to analyze very large healthcare datasets. Health Care Manag Sci 7:163–171
Stimson DH, Stimson RH (1972) Operations research in hospitals: diagnosis and prognosis. Hospital Research and Educational Trust, Chicago
Sobolev B, Harel D, Vasilakis C, Levy A (2008) Using the Statecharts paradigm for simulation of patient flow in surgical care. Health Care Manag Sci 11:79–86
Stummer C, Doerner K, Focke A, Heidenberger K (2004) Determining location and size of medical departments in a hospital network: a multiobjective decision support approach. Health Care Manag Sci 7:63–71
Sun L, DePuy GW, Evans GW (2014) Multi-objective optimization models for patient allocation during a pandemic influenza outbreak. Comput Oper Res 51:350–359
Sundaramoorthi D, Chen VCP, Rosenberger JM, Kim SB, Buckley-Behan DF (2009) A data-integrated simulation model to evaluate nurse–patient assignments. Health Care Manag Sci 12:252–268
Sundaramoorthi D, Chen VCP, Rosenberger JM, Kim SB, Buckley-Behan DF (2010) A data-integrated simulation-based optimization for assigning nurses to patient admissions. Health Care Manag Sci 13:210–221
Swisher JR, Jacobson SH (2002) Evaluating the design of a family practice healthcare clinic using discrete-event simulation. Health Care Manag Sci 5:75–88

Swisher JR, Jacobson SH, Jun JB, Balci O (2001) Modeling and analyzing a physician clinic environment using discrete-event (visual) simulation. Comput Oper Res 28:105–125
Testi A, Tanfani E, Torre G (2007) A three-phase approach for operating theatre schedules. Health Care Manag Sci 10:163–172
Testi A, Tanfani E (2009) Tactical and operational decisions for operating room planning: efficiency and welfare implications. Health Care Manag Sci 12:363–373
Thokala P, Duenas A (2012) Multiple criteria decision analysis for health technology assessment. Value Health 15:1172–1181
Tiwari V, Heese HS (2009) Specialization and competition in healthcare delivery networks. Health Care Manag Sci 12:306–324
Toro-Díaz H, Mayorga ME, Chanta S, McLay LA (2013) Joint location and dispatching decisions for emergency medical services. Comput Ind Eng 64:917–928
Utley M, Jit M, Gallivan S (2008) Restructuring routine elective services to reduce overall capacity requirements within a local health economy. Health Care Manag Sci 11:240–247
van de Klundert J, Muls P, Schadd M (2008) Optimizing sterilization logistics in hospitals. Health Care Manag Sci 11:23–33
VanBerkel PT, Blake JT (2007) A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. Health Care Manag Sci 10:373–385
Varela PS, de Andrade MG, Fávero LPL (2010) Production efficiency and financing of public health: an analysis of small municipalities in the state of São Paulo. Health Care Manag Sci 13:112–123
Vasilakis C, El-Darzi E (2001) A simulation study of the winter bed crisis. Health Care Manag Sci 4:31–36
Verter V, Lapierre SH (2002) Location of preventive health care facilities. Ann Oper Res 110:123–132
Vos L, Groothuis S, van Merode GG (2007) Evaluating hospital design from an operations management perspective. Health Care Manag Sci 10:357–364
Xu G-C, Zheng J, Zhou Z-J, Zhou C-K, Zhao Y (2015) Comparative study of three commonly used methods for hospital efficiency analysis in Beijing Tertiary Public Hospitals, China. Chin Med J 128(23):3185–3190
Yaesoubi R, Roberts SD (2010) A game-theoretic framework for estimating a health purchaser's willingness-to-pay for health and for expansion. Health Care Manag Sci 13:358–377
Zhang Q, Huang Y (2012) Early warning index selection and weight assignment for city significant emergency in uncertain environment. Adv Inf Sci Serv Sci 4(23):490–497
Zhang Y, Berman O, Marcotte P, Verter V (2010) A bilevel model for preventive healthcare facility network design with congestion. IIE Trans 42:865–880
Zeng Q-L, Li D-D, Yang Y-B (2013) VIKOR method with enhanced accuracy for multiple criteria decision making in healthcare management. J Med Syst 37:9908
Zonderland ME, Boucherie RJ, Litvak N, Vleggeert-Lankamp CLAM (2010) Planning and scheduling of semi-urgent surgeries. Health Care Manag Sci 13:256–267

Chapter 2
A Comprehensive Survey on Healthcare Management

Sezi Cevik Onar, Basar Oztaysi, and Cengiz Kahraman

2.1 Introduction

Healthcare management is the scientific field which provides leadership and direction to healthcare organizations. The public health systems, healthcare systems, hospitals, and hospital networks are the main research areas in healthcare management. In these healthcare organizations, there usually exist numerous types of healthcare problems which might be micro such as scheduling of an operating rooms or macro such as layout of the whole healthcare organization. These problems can be solved by quantitative and qualitative techniques such as operations research techniques, simulation or decision making techniques.

Management of healthcare organizations needs complex and dynamic operations. Managers in these organizations should perform a successful leadership, supervision, and coordination of the employees. They should be aware of the quantitative and qualitative techniques that can be used in the solutions of various problems of healthcare organizations. In the literature, there are many healthcare applications of these techniques. However, there is a need for a work classifying these studies based on the types of techniques and the healthcare problems solved by these techniques.

Figure 2.1 illustrates the increasing trend of healthcare management publications over the years. The statistics in Fig. 2.1 are from Scopus database with the keyword *healthcare management* in the title of the publications. This search gave us 2,250 papers. There is a significant increase in the number of publications after 1998. This proves that a significant importance is given to healthcare management research in the literature for the last two decades.

S. Cevik Onar (✉) • B. Oztaysi • C. Kahraman
Department of Industrial Engineering, Istanbul Technical University, Macka, Istanbul, Turkey
e-mail: sezi@outlook.com

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_2

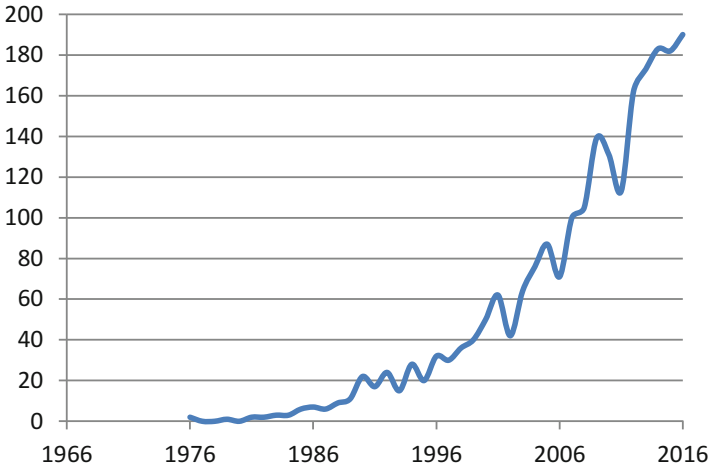

Fig. 2.1 Healthcare Management publications with respect to years

The aim of this chapter is to exhibit the existing position of healthcare management in the literature. We classified the healthcare problems and their solution techniques in the literature in order to see the general frame of healthcare management. OR techniques are also included within this classification. This aims at showing the place of OR in the whole picture. A comprehensive and up to date literature survey is conducted by using a broader perspective that considers both qualitative and quantitative techniques.

The rest of the chapter is organized as follows. Section 2.2 presents the literature review results based on the classifications with respect to papers conducting a review of qualitative and quantitative techniques on HCM. Section 2.3 classifies the techniques used in healthcare management. Finally, Sect. 2.4 concludes the chapter and presents future directions.

2.2 Sub-classifications of HCM Studies

There is a significant increase in the HCM studies in the recent years. United States and United Kingdom are the first two by far leading countries in publishing healthcare works. In this manner, the second class of countries includes Finland, Canada, India, Italy, France, and Germany; the third class includes Brazil, Australia, Israel, Netherlands, and Taiwan. Figure 2.2 illustrates the countries most publishing healthcare works. It is interesting that China, one of the leading countries in many research areas, is about 20th place in this list. A publication is counted more than one if its authors are from different countries.

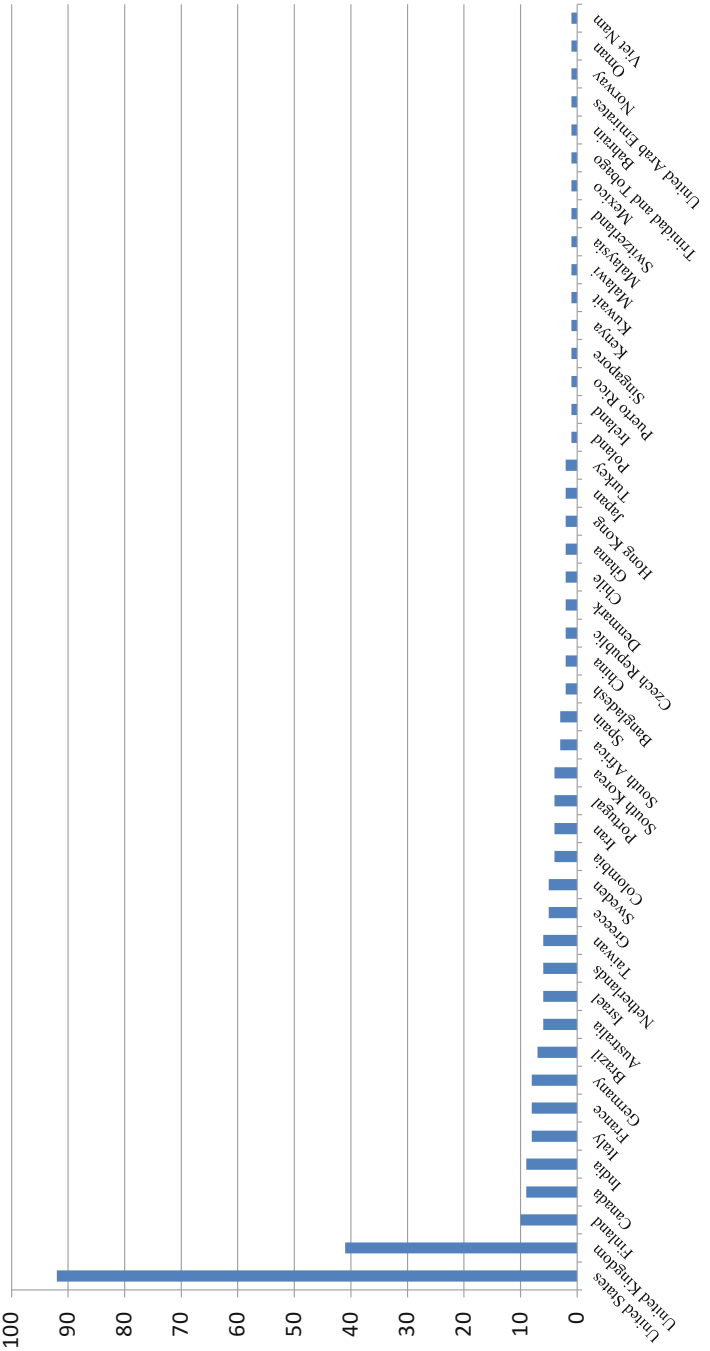

Fig. 2.2 Countries publishing in Healthcare Management

Some institutes focus on healthcare research more than the others. Texas A&M University, University of Southhampton, Ben-Grurion University of the Negev, Cardiff University, and VA Medical Center are the leading institutes publishing healthcare papers.

In the following, HCM papers are classified as literature review papers, qualitative and quantitative techniques, case studies, and performance measurement paper, and some representative papers are given.

2.2.1 HCM Papers Making Literature Review

Some researchers have attempted to review the literature on qualitative and quantitative techniques utilized in healthcare management systems. Fakhimi and Mustafee (2012) focus on synthesizing extant literature in healthcare OR by classifying papers based on OR techniques, application category, healthcare specialty, among others. The scope of this review article is limited to OR studies undertaken in the UK. Ozcan (2009) presents the quantitative methods used in healthcare management with their applications, including decision trees, forecasting techniques, facility layout and facility location techniques, and scheduling techniques. Tanfani and Testi (2012) edit a book on advanced decision making methods applied to healthcare, including the applications of operational research, statistical and economic decision-making tools in the field of healthcare delivery. Cox (2006) presents the traditional quantitative risk assessment techniques for the human health consequences of using antibiotics in food animals. These techniques include Bayesian Monte Carlo analysis, rapid risk rating technique, etc.

Some other researchers made literature reviews on various healthcare problem areas. Lavis et al. (2005) review studies of decision-making by health care managers and policy-makers. They claim that literature reviews are necessary for developing better inform health care management systems and effective policy-making. Kontio et al. (2007) present a literature review on the benefits of healthcare information systems and enterprise resource planning (ERP) systems in healthcare. The study is conducted using a meta-summary technique for qualitative research. Kamarudeen (2010) reviews the literature to reveal the impact of healthcare system on the amenable mortality rates in the OECD countries. Rais and Vianaa (2011) review key contributions addressing contemporary optimization issues in the domain of healthcare. The study focusses on the optimization problems in current research activities and the solution techniques used for solving the optimization problems. Wang et al. (2011) present a literature review and analyze papers which use industrial and systems engineering and operations management methods to improve psychological healthcare. Matopoulos and Michailidou (2013) investigate the collaborative practices in the healthcare supply chain and give insights into hospital-vendor operations. Ashrafi et al. (2013) conduct a systematic review in order to reveal decision support applications and their effects on healthcare. The results indicate that decision support systems are applied to the five main areas, namely,

disease progress management, care and treatment, drug prescribing, evaluation and prevention. Fakhimi and Probert (2013) review the operations research techniques that are used in healthcare, and categorize these studies based on the application type and operation research technique employed. The results indicate that the majority of studies focus on simulation. Dobrzykowski et al. (2014) conduct a structured literature review on operation management and supply chain management (SCM) studies for healthcare. Al-Balushi et al. (2014) conduct a literature review in order to define the readiness factors that are critical to the application and success of lean operating principles in healthcare organizations. The above literature reviews focus on best practices and specific techniques and areas such as ERP usage and SCM practices in healthcare management. Their focus is not directly the quantitative and qualitative techniques used in healthcare management.

2.2.2 Quantitative and Qualitative Techniques in HCM

Since the 1960s, quantitative techniques and mathematical models, such as operations research models, have been applied to a range of healthcare problems. However, OR techniques have not been frequently used in the solution of healthcare problems by clinicians, health managers and policy-makers. Brailsford (2005) focuses on simulation models in healthcare management and briefly describes one successful implementation and suggests some potential ways forward. Green (2012) describes the essential features and critical issues of the United States (US) healthcare system that provide opportunities for operations researchers to make significant contributions. Garg et al. (2012) develop an intelligent patient management and resource planning model for complex, heterogeneous, and stochastic healthcare systems. Motamarri et al. (2014) analyze patients' perceptions on mobile healthcare services by multiple discriminant analysis. Sadatsafavi et al. (2015) investigate employees' perceptions of healthcare facilities and differences across demographic groups by using principal component analysis, confirmatory factor analysis, and invariance analysis. Xie and Lawley (2015) indicate that innovative OR techniques are developed for operating room planning, emergency department staffing, breast cancer screening, radiotherapy treatment planning, home healthcare planning, long-term care planning and scheduling.

In the following, we present some statistics based on the operations research applications in healthcare management. The data are obtained from Scopus database with the inputs healthcare and operations research. The distribution of the healthcare publications in OR is composed of articles with a percentage of 55; conference papers with a percentage of 31; review papers with a percentage of 7; book chapters with a percentage of 4; and books, notes, and editorials with a percentage less than 3. The main research areas of the operations research based healthcare publications are medicine, engineering, computer science, business and management, decision sciences, social sciences, health professions, mathematics, and nursing, respectively.

Qualitative methods for the solutions of healthcare management problems have been employed in some works. Robert et al. (1999) use Delphi study for identifying new healthcare technologies. Zonca et al. (2015) specify the competition in healthcare with a focus on surgery in the Czech Republic and use Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis.

2.2.3 Case Studies in HCM

Case studies in healthcare management have been also handled in numerous papers. Lavy and Shohet (2009) develop an integrated facilities management (FM) model for healthcare facilities and investigate the effectiveness of the model in terms of maintenance and performance management in a real world case study. Krey et al. (2010) provide an overview of the interrelated information technology (IT) governance frameworks and best practice models, analyze the potential impacts of IT governance on the Swiss healthcare and give an outlook of the related future research. Foster et al. (2010) illustrate the use and value of the tools of operations research in healthcare and focus on queuing theory. Applying queuing theory in a hypothetical drug treatment facility, some of the key performance measures, such as average waiting time for admission, are modeled using mathematical expressions. McAlearney et al. (2010) focus on potential links between high-performance work practices and quality of care and patient safety in US healthcare organizations. After an extended literature review the authors generate a model and confirm the model by five case studies. Nemeth and Cook (2010) focus on Resilience engineering in healthcare enterprises. Based on a 5-year case study, the authors present a concept for an infusion device interface that would contribute to resilience. Ahsan et al. (2010) design and provide the insight of an Enterprise architecture approach to process architecture for healthcare-IT alignment. They analyze healthcare organizational processes using a specific case study and conceptualize this analysis in order to provide an overview of healthcare processes in the context of enterprise architecture to improve healthcare management. Van Vactor (2011) presents a collaborative communications model which provides information to healthcare supply chain managers and administrators. Using the data obtained from a case study including healthcare supply chain managers in the US Army, the effects of collaborative communications on healthcare supply chain management is pointed out. Bora et al. (2011) describe a methodology to support the evaluation of the benefits provided by Radio Frequency Identification (RFID) on product traceability applications in healthcare sector and provide a real world case study. Bullock et al. (2013) analyze a United Kingdom (UK) knowledge exchange program designed to bring together healthcare managers and researchers with a case study.

2.2.4 Performance Measurement in HCM

There are several quantitative and qualitative techniques aiming at measuring the performance level of a healthcare organization or comparing its relative position with the other organizations. A healthcare organization should measure its performance since measuring performance provides quality improvement, transparency, accreditation, recognition as a patient centered medical home, and participation in financial incentive programs or demonstrations. Literature review reveals the common performance measures for healthcare organizations as in Table 2.1.

A performance measurement system should take care of all the indicators given in Table 2.1. The performance of healthcare organizations can be compared based on a multicriteria decision making (MCDM) technique which can consider all the criteria given above. Some examples of MCDM techniques are analytic hierarchy process (AHP), simple additive weighting (SAW), Preference Ranking Organisation Method for Enrichment Evaluations (PROMETHEE), Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS), ELimination Et Choix Traduisant la REalité (ELECTRE), or VIseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR).

2.3 Classification of Techniques Used in Healthcare Management

In this section, we classify HCM techniques as follows: Simulation, multicriteria decision making, mathematical programming, statistical decision making, data envelopment analysis (DEA), data mining, engineering economics, human factors engineering, structural equation model, design of experiments, system dynamics, case study, qualitative research and other approaches. Figure 2.3 illustrates the literature review results on qualitative techniques used in HCM research. This literature review is based on the search conducted in Scopus database. The keywords used are the name of the techniques and *healthcare*, which are searched in the title of the publications. The related studies under each technique are briefly given in the following.

2.3.1 Simulation

The objective of simulation is to imitate the real-world processes or systems over time, and it is based on repeated trials. A simulation model can quickly investigate the effects of a change in a real life situation that may take place over several years. It can be used to study complex systems like healthcare systems that would

Table 2.1 Common performance measures for healthcare organizations

Quality of care	Utilization-cost-efficiency	Satisfaction- reports of care	Financial performance	Others
Hospital-level mortality, complication, and infection rates	Cost per adjusted discharge	Patient-reported satisfaction	Operating revenues	Availability of foreign-language written materials
Rates of specific medical errors or other patient safety issues	Bed occupancy rate	Communication/information and consistent messages from multiple providers and wait times and ease of access	Operating expenses	Availability and ease of use of translation services
Five-year survival rates for specific cancers	Length of stay	Appearance of facilities and parking/food/other services	Pharmacy cost	Number and scope of cultural competence training program
Low birth-weight or pre-term birth rate	Patients per physician per day	Control of pain or other symptoms	Total cost per patient	Uncompensated care
Unexpected return to surgery	Nurses per service	Coordination of care	Medical cost per medical visit	Care provided in public programs
Reducing variability in clinical care	Physician per service and Patients per service per day	Respect for values and preferences	Long term debt to equity ratio and Change in net assets to expense ratio	Numbers served in free clinical service programs

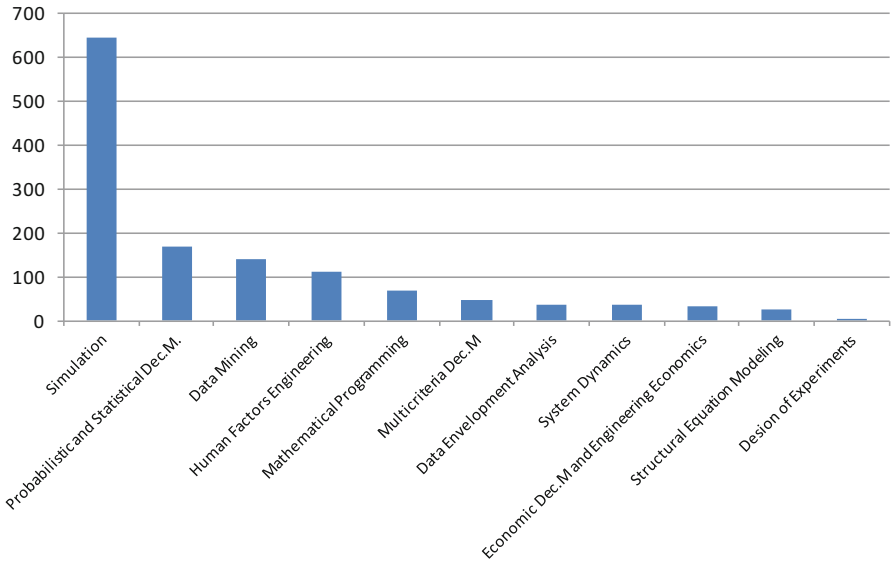

Fig. 2.3 Frequencies of qualitative techniques used in HCM literature

otherwise be difficult to investigate. However, simulation may not well model a complex multidimensional system if we do not have sufficient data to produce a mathematical model.

Simulation in healthcare is often used in safety and quality-oriented training programs, development of educational and competency assessment standards, virtual reality, epidemiologic modeling, and molecular, pharmacologic, and disease modeling. Simulation in healthcare provides a range of readily available learning opportunities, and the freedom to make mistakes and to learn from them. The learning experience can be customized, and thus, complex procedures and rare diseases can be tried by simulation training methods. Simulation also provides a detailed feedback and evaluation.

In the following, we summarize the literature on simulation-based solutions of healthcare problems.

Ramakrishnan et al. (2004) present the results of a collaborative research effort with a healthcare provider for a digital image archiving system within its radiology services. The objective is to maximize patient throughput and minimize report generation time. They build a simulation model to evaluate the different scenarios expected to 'optimize' the response variables. Lavy and Shohet (2007b) develop a model to adjust the allocation of maintenance resources to prevailing service conditions in healthcare facilities. The configurations integrating occupancy and environmental conditions are investigated through simulations and compared to a reference configuration. Sachdeva et al. (2007) try to combine OR methodologies

to achieve greater acceptance of results for an organizational change. Patient flow is modeled using simulation. Results from simulation, particularly for politically sensitive issues, are persuasive but inadequate to result in change. Eldabi et al. (2007) use simulation within healthcare settings based on the survey data obtained from the experts composed of academics and industrialists, a critical analysis is applied to find the differences between what exists and what could be created based on outlining some major themes. Chandra (2008) presents a healthcare supply chain template utilizing e-commerce strategy. By using simulation, optimization and information-sharing techniques are used to optimize purchasing and inventory policies. Gonsalves and Itoh (2009) propose a simulation model that contains both the subjective and objective elements in the patients' evaluation of healthcare services. The proposed model is simulated via discrete event simulation, and Genetic Algorithm is used to optimize the model. Brailsford et al. (2010) focus on the idea of combining discrete-event simulation and system dynamics and describe two practical healthcare examples of combined discrete-event simulation and system dynamics models. Gaion et al. (2009) model and simulate an alarm management system in a Colored Timed Petri Net framework to be used for testing scenarios of alarm management of healthcare devices with different levels of workload and resources. Cabrera et al. (2012) present an Agent-Based modeling simulation to design a decision support system for Healthcare Emergency Department (ED). The objective of the proposed procedure is to optimize the performance of such complex and dynamic Healthcare EDs by optimal staff configuration including doctors, triage nurses, and admission personnel. Knight et al. (2012) propose a simulation model to study the effects of patient behavior using discrete event simulation. Instead of utilizing some probability distribution for individual decisions, a decision making model is used based on system observations. The proposed model can be used for decision makers to improve overall system performance and for solving location-allocation problems. Robinson et al. (2012) utilize simulation in the implementation of lean in healthcare. The study uses both the impact of discrete-event simulation and lean approaches to the improvement of healthcare systems. Buyurgan and Farrokhvar (2015) develop a simulation model to investigate adverse events and patient safety in healthcare due to poor supply chain management practices, and inadequate and disorganized product validation procedures and compare different scenarios for patient safety, care delay, and system efficiency. Baril et al. (2016) aim at allowing a rapid and successful implementation of the solutions developed during the Kaizen. They develop a discrete event simulation to test scenarios defined by team members during a Kaizen event. Before Kaizen, some problems with the system are the limited treatment room capacities in the mornings and computerized appointment system that can schedule one treatment at the same time. Applying Kaizen, patient delays before receiving their treatment are reduced by 74% after 19 weeks.

2.3.2 Multicriteria Decision Making

MCDM is the optimum decision making process under the existence of various and conflicting criteria. MCDM is divided into two subgroups called multi-attribute decision making (MADM) and multi-objective decision making (MODM). MADM includes the techniques composed of discrete decision spaces whereas MODM includes the techniques composed of continuous decision spaces. MCDM methods are widely used in public and private sector decisions on transport, education, investment, environment, energy, defense, and so forth.

The health care industry has been relatively slow to apply MCDA. But as more healthcare researchers and practitioners have become aware of the techniques, there has been a sharp increase in its healthcare applications. Decision criteria must be weighted, and goal achievement must be scored for all alternatives. Methods of multi-criteria decision making are available to analyze and appraise multiple clinical endpoints and structure complex decision problems in healthcare decision making.

Bose (2003) presents and describes the knowledge management capabilities, the technical infrastructure, and the decision support architecture for such an HCM system and provides a decision support infrastructure for clinical and administrative decision-making. Aktas et al. (2007) propose a management-oriented decision support model to assist health system managers in improving the efficiency of their systems. The conditional dependencies and uncertainties are represented by using Bayesian Belief Networks. Zeng et al. (2013) propose an improved VIKOR method with enhanced accuracy to support the decision making in healthcare management. Dehe and Bamford (2015) compare two multiple criteria decision analysis models for a healthcare infrastructure location decision. Evidential Reasoning (ER) is used to solve the model, and Analytical Hierarchy Process (AHP) is used to compare the processes and results. Hussain et al. (2016) develop a framework to assist lean deployment in Abu Dhabi public healthcare delivery systems. Twenty one healthcare wastes are ranked based on the evaluations of local situations by experienced healthcare professionals. Marcarelli (2016) evaluate healthcare policies with benefit/cost analysis by the analytic hierarchy process. The effectiveness and efficiency of some policies and their costs are considered in the evaluation process.

2.3.3 Mathematical Programming

Mathematical programming requires the use of a computer program to make an optimum decision. It includes probability and mathematical models to solve the problems. It is one of several OR techniques whose particular characteristic is that the best solution to a model is found by optimization software. Mathematical programming enables simultaneous consideration of multiple constraints and sensitivity analysis and provides an efficient tool for healthcare professional (Earnshaw and Dennett 2003).

Rönnberg and Larsson (2010) aim at developing an optimization tool that automatically delivers a usable schedule based on the schedules proposed by the nurses. The authors develop a mathematical model for a typical Swedish nursing ward and analyze the results. Adasme et al. (2015) develop a minmax robust formulation for routing in healthcare wireless body area networks. The proposed formulation minimizes the highest power consumption of each biosensor node placed in the body of a patient subject to flow rate and network topology constraints. The formulation includes an equivalent polynomial formulation of the spanning tree polytope to avoid having an exponential number of cycle elimination constraints and a mixed integer linear programming (MILP) formulation of the traveling salesman problem. Paschou et al. (2015) develop a personnel rostering system for healthcare units, which incorporate mobile technologies. This system minimizes the time and other bureaucratic delays in personnel scheduling.

Operating room planning is another area that OR techniques are often used. Landa et al. (2016) deal with the operating room planning problem at an operational planning level. The problem addressed consists of two interrelated subproblems usually referred to as "advance scheduling" and "allocation scheduling." In the first sub-problem, the decisions considered are the assignment of a surgery date and an OR block to a set of patients to be operated on over a given planning horizon. The second aims at determining the sequence of selected patients in each OR and day. Roshanaei et al. (2017) aim at selecting patients with the highest priority scores and schedule them in the current planning horizon. They determine the number of surgical suites and operating rooms required to accommodate the schedule at minimum cost.

A focus in global healthcare today is on the task of improving inventory management. The main challenges in this task include uncertainty in demand and limited human resources.

Danas et al. (2006) identify the inefficiencies of the logistics systems of Greek hospitals through the management of medicine stock within the hospital pharmacy. Fortsch and Khapalova (2016) address the challenges faced by blood centers by introducing practical methods for accurate blood demand forecasting, which will allow for lowering of costs, reduction of blood wastage, and conservation of limited resources. In all locations, demand-forecasting is completed using the popular Microsoft Excel spreadsheet software; however, this field research study shows the demand for blood is non-stationary and cannot be accurately forecasted using Excel, at least not without writing a macro. Hence, they use multiple approaches to predict blood demand. At the end of the study, the Box-Jenkins methodology is shown to be the optimal choice to forecast demand. Saedi et al. (2016) develop a stochastic model to find the optimal inventory policy for a healthcare facility. They proactively minimize the effects of drug shortages under uncertain disruptions and demand.

2.3.4 Probabilistic and Statistical Decision Making

Statistical decision making is the process of analyzing data and using methods of statistical inference in making business decisions. Both probabilistic and statistical techniques such as statistical sampling and sampling distributions, point estimation and confidence intervals, hypothesis testing, correlations among variables and multivariate analysis are used in this process.

The capability of making statistical analyses has an extreme importance in healthcare management. Healthcare professionals generally use empirical information via statistical summaries to make decisions rather than deep data analyses. If statistical education of healthcare professionals falls short, finding the possible relations among diseases, estimation of parameters, making inferences about these parameters, multivariate analyses including clustering, factor analysis, etc. may not be possible.

Guo et al. (2008) develop a community healthcare competency scale for public health nurses (PHNs). They explore community healthcare competency of PHNs in Taiwan by using a cross-sectional research design to collect data. Chandrasekaran et al. (2012) investigate the effect of process management on clinical and experiential quality. The data gathered at various time intervals is used to statistically test the developed hypotheses. Four important implications emerge from this work. Uddin et al. (2012) measure the effectiveness of static clinic and satellite clinics to provide primary healthcare services to street-dwellers. Data collected before and after the implementation of the clinics are compared with a t-test. Vozikis et al. (2012) propose a specialized partially observable Markov Decision Process form in order to determine an optimal or nearly optimal policy for the treatment of patients with ischemic heart disease. The proposed approach has a practical advantage over clinical studies such as no risk for the life of patients and low cost. Weidmer et al. (2014) develop a system for the consumer assessment of healthcare providers for in-center hemodialysis patients. The reliability and validity of the survey are assessed with statistical methods. Van Minh et al. (2014) evaluate the primary healthcare system capacities for responding to storm and flood-related health problems by using self-administered questionnaires, in-depth interviews and focus groups discussions. Dobrzykowski and Tarafdar (2015) develop and test research hypotheses linking information Technologies. They use a paired sample of primary survey data and secondary archival data for 173 hospitals in the US. They find that increased information exchange relationship drives provider-patient communication, and increased social interaction ties drive information exchange relationship. Michailidou et al. (2015) compare the hospital charges accrued following appendectomies operations in the pediatric population. A total of 264 cases from 2007 to 2013 are reviewed, and the results indicated higher costs at laparoscopic operations. Fervers et al. (2015) examine the effects of globalization on healthcare expenditure. The research problem is whether the relationship changes

with respect to the types of healthcare systems. They analyze 22 OECD states between 1980 and 2009 in pooled time-series regressions. They find that an increase in economic openness leads to lower spending growth and to stronger in countries with social health insurance systems.

2.3.5 Data Envelopment Analysis (DEA)

Data envelopment analysis is a linear programming based technique for measuring the relative efficiencies of decision making units by employing multiple inputs and outputs. Two types of scales are used in DEA: constant returns to scale (CRS), and variable returns to scale (VRS). In CRS approach, the output changes by the same proportion as inputs are changed. In VRS approach, production technology may exhibit increasing, constant and decreasing returns to scale.

Malhotra et al. (2015) use data envelopment analysis to benchmark the performance of 12 publicly managed care organizations against one another for the period 2009–2011. They find that only six companies out of 12 are 100% efficient. They also identify the areas in which inefficient companies are lagging behind their efficient peers. Davey et al. (2015) use DEA technique by using the constant ratio to scale to compare four decision making units for the efficiencies of two private health centers of a private medical college and two public health centers. DEA technique reveals that the government health facilities group are more efficient in the delivery of primary healthcare services with respect to private training health facilities group.

2.3.6 Data Mining (DM)

Data mining is the process of analyzing and discovering patterns in data by using different perspectives such as artificial intelligence, machine learning, statistics, and database systems. It aims at extracting information and transforming data into an understandable structure for further use.

DEA enables using multiple inputs such as the number of beds, doctors and nurses and outputs such as patient days and total immunization for measuring the performance of healthcare systems. In some of DEA applications, the weights of inputs and outputs can be defined (Al-Shayea 2011).

Huang et al. (1995) propose an agent-based system in order to help manage the care process in real-world settings by combining artificial intelligent and agent techniques. Swangnetr et al. (2010) compare two simulated robot medicine delivery experiments with different participant age groups and robot configurations by using a meta-analysis of data and statistical and machine learning methods. Utter et al. (2010) conduct a retrospective cross-sectional study including 11 indicators from 18 geographically different academic medical centers based on the medical records using a standard instrument and descriptive analysis. Darrel et al. (2014) present a

literature review on the quantifiable and measurable benefits of big data analytics in healthcare systems. The main advantages of big data analytics usage are obtained in the improved outcomes for patients and lower costs for healthcare providers. Ramírez-Ríos et al. (2015) examine DM algorithms as a feasible and necessary strategy for optimal management of databases (DB) in the national healthcare systems. They deal with the management of multiple DB that considers patient's affiliation information under the supervision of the authorities in healthcare. Their DM analysis detects frauds and other type of duplicities.

2.3.7 Economic Decision Making and Engineering Economics

Engineering economics is the science of giving economic decisions based on discounted cash flow techniques such as present worth analysis, annual worth analysis, the rate of return analysis, etc. Time, cash, and interest are the three parameters of engineering economics. The economic analysis methods can also be used under risk and uncertainty conditions.

Through better healthcare system design and capital investments in new health technologies, such as electronic medical records, telemedicine, and imaging systems, healthcare managers can manage disbursements and benefits. The time value of money can be considered through engineering economics science, and thus correct economic decisions can be made. Sensitivity analysis should be conducted before giving risky investment decisions since health technologies are too expensive. Multiparameter sensitivity analysis rather than one-at-a-time sensitivity analysis should be applied to analyze the effects of changes in two or more parameters.

Johnson (2008) develops some spreadsheet functions that enable queuing theory in healthcare systems. The proposed functions can be used for better process understanding leading to better decision making and optimization of the healthcare budget. Dortland et al. (2013) reveal the position of real estate departments through an exploratory survey among health organizations. They consider the type of project coalitions and the rationale behind this choice and the flexibility regarding a real option. Devaraj et al. (2013) examine information technology investments among hospitals and how it influences patient care and financial performance with an operations management-based perspective on the effect of IT in streamlining hospital operations. Zadeh et al. (2015) describe a framework that can facilitate the implementation of evidence-based design (EBD), and clarify the related safety and quality outcomes for the stakeholders. They use engineering economy tools including present values, internal rates of return, and payback periods to evaluate the return on investments, in which facility design and operation interventions resulted in reductions in hospital-acquired infections, patient falls, staff injuries, and patient anxiety.

2.3.8 Human Factors Engineering

Human-factors engineering is the science that deals with the application of information on physical and psychological characteristics to the design of devices and systems for human use. In order to provide safety, effectiveness, and ease of use, human factor engineering focuses on both human strengths and limitations in the design of interactive systems.

Human factors engineering has recently been used in healthcare management to increase reliability, especially in the operating room. Human factors engineers test new systems and equipment under real-world conditions in order to identify potential problems and unintended consequences of new technology. These usability tests can be applied to various healthcare management problems such as data and information processing.

Boston-Fleischhauer (2008) uses human factors engineering and reliability science for enhancing existing operational and clinical process design methods in healthcare. Van De Weerdt and Baratta (2012) focus on analyzing working conditions of home healthcare services, which is a growing service area in Europe, such as aides and nurses. The authors analyze the impacts of home healthcare works regarding job satisfaction, well-being, emotions at work, relationships with the others and occupational stress.

2.3.9 Structural Equation Modeling (SEM)

Structural equation modeling is composed of a diverse set of mathematical models, computer algorithms, and statistical methods such as confirmatory factor analysis, path analysis, and partial least squares path analysis that fits networks of constructs to data. SEM can deal with measurement error, enables examining, and modelling complex healthcare problems (Beran and Violato 2010).

Gowen et al. (2006) examine healthcare quality program practices, employee commitment, and control initiatives, and perceived results based on the responses from Quality and Risk Directors of 372 U.S. hospitals. The results of SEM indicate that both quantitative and qualitative quality program results are mainly related with employee commitment and control initiatives. Haggerty et al. (2011) compare how well accessibility is measured in four different subscales that evaluate primary healthcare from the patient's perspective based on SEM analysis. They use the results of a survey of 645 adults with at least one healthcare contact in the previous 12 months.

2.3.10 Design of Experiments (DOE)

The design of experiments deals with the determination of the relationship between factors affecting a process and the output of that process. The primary purpose is to find cause-and-effect relationships between the process inputs and outputs. DOE techniques are commonly used in clinical evaluations but their usage in the production and design phases are limited. Designing effective experiments increases the reliability and efficiency of healthcare management.

Ramakrishnan et al. (2005) focus on identifying appropriate modifications to the existing workflow at the computed tomography (CT) scan area of a healthcare provider while transitioning from a film-based image archiving system to a digital system. Apart from the workflow, the flow of information also needs to be modified and streamlined. The ultimate goal is to maximize patient throughput and minimize report generation time. Industrial engineering tools such as process mapping and time study are used to understand the initial flow of operations. A Design of Experiments based approach is used to identify the effect of the variables in the system and the interaction amongst them. Modeling and simulation are used to analyze and quantify the potential benefits that can result from the implementation of the digital image archiving system. Savsar and Al-Ajmi (2012) determine the significant factors that cause delays in surgery operations and affect the productivity of surgery clinics in a hospital. Based on data collected from surgery clinics of an international hospital, design of experiment approach is used to determine the significance of the effects of the factors.

2.3.11 System Dynamics

System dynamics aims at understanding the nonlinear behavior of complex systems over time based on flows, internal feedback loops, and time delays. It is a computer-aided approach applied to dynamic problems arising in complex social, managerial, economic, or ecological systems. It first defines problems dynamically and proceeds through mapping and modeling stages. System dynamics enables modeling complex healthcare problems such as multiple interacting diseases, supply chain problems, and expansion of diseases (Homer and Hirsch 2006).

Lane et al. (2000) develop a system dynamics model of the interaction of demand pattern, resource deployment, hospital processes and bed numbers. They find that some delays to patients are unavoidable and reductions in bed numbers do not increase waiting times for emergency admissions. Brailsford (2008) illustrates several examples of system dynamics in healthcare organizations and discusses the possible reasons for the popularity of system dynamics for healthcare modeling. Samuel et al. (2010) analyze health service supply chain systems using system dynamics, where three service stages are presented sequentially. Zamora Aguas et al. (2013) develop a system dynamics model in order to assess supply risk impact

in the oncological medicine supply chain in Colombia. Supply networks, supply chain costs, improving service, quality and opportunity performance indexes are included into the model. Rich and Piercy (2013) develop a systems dynamics model of hospital healthcare in order to capture the problems in the existing system and their inter-relationships.

2.3.12 Qualitative Approaches

Qualitative methods are used to gain an understanding of primary reasons, opinions, and motivations about research problems. Qualitative research is also used to uncover trends in thought and ideas, and exhibit details about the problem. Some standard qualitative approaches include focus groups, individual interviews, and participation/observations.

Qualitative approaches are relatively rarely used in healthcare management but they add significant value to healthcare management. Especially, social and cultural aspects of healthcare management can be revealed with qualitative research. Qualitative approaches provide important insights into health-related phenomena, generating new ways for empirical questions. Qualitative research is based on health related lived experiences and relational processes as the basis of social phenomena.

Finstuen and Mangelsdorff (2006) identify the mentoring and executive competencies required among preceptors of a graduate program in health and business administration. They specify the requisite skills, knowledge, and abilities needed to achieve those competencies by using a Delphi methodology through e-mails. Hunt (2009) focus on relief operations and aims at exploring the moral experience of Canadian healthcare professionals during humanitarian relief work. The authors conduct 18 semi-structured individual interviews based on Interpretive Description methodology. Bauernschmitt and Conradie (2010) establish a descriptive research to present to what extent private healthcare providers have contemporary knowledge and understanding of supply chain practices and what extent these providers adopt and apply such knowledge to recognized practices and concepts. Farinella et al. (2011) describe the outcomes of a case study on the regional differences in implementation of "stroke networks" in Italy, which is one of the most important health issues in Italy. The results of 52 in-depth interviews and six focus groups indicate that early diagnosis, delivery of treatment and rehabilitation therapy can reduce the risks of death and disability. Papadopoulos et al. (2011) conduct a qualitative study in order to explore the dynamics in the implementation of a process improvement methodology using actor-network theory. They illustrate the utility of actor-network theory in articulating the dynamic nature of networks underpinning socio-technical change. Nelson (2011) presents a descriptive study and describes the perceptions of staffing adequacy of healthcare team members working together after conducting semistructured interviews in a cancer center. Hadziabdic et al. (2011) focus on exploring the problems reported by healthcare professionals in primary healthcare concerning the use of interpreters and what the problems lead

to. The authors use qualitative content analysis of 60 real-life incident reports to find out the major problems. Jaafaripooyan et al. (2011) identify performance measures to evaluate accreditation programs in healthcare based on qualitative methods, including snowball sampling technique, email interviews, and thematic content analysis.

2.3.13 Other Approaches

Some healthcare management papers cannot be classified with respect to the above-mentioned approaches. We categorize these works as follows:

2.3.13.1 Cost Analysis in Healthcare

Nicholson et al. (2004) focus on inventory costs and service levels in a healthcare organization. The authors compare two models, three-echelon distribution network managed by the healthcare organization, and two-echelon distribution network outsourced to a third party, for non-critical inventory items. The results indicate that outsourcing distribution of non-critical medical supplies reduce inventory cost savings while service levels are not changed.

2.3.13.2 Quality in Healthcare

Dey and Hariharan (2006) propose a model for identifying problems and evaluating the performance of healthcare services. The authors apply the model, logical framework analysis in three services of a hospital and show the effectiveness of the proposed method with the case study. Mohammadi et al. (2007) suggest a quality based outline for a customer-driven health system. The authors determine six types of customers, nine types of outputs and the various operations associated with them. De Mast et al. (2011) focus on healthcare processes in order to provide a unifying and quantitative framework. In their proposed methodology, the authors integrate the various process improvement approaches such as six sigma, lean thinking, and total quality management in order to provide conceptual models and practical templates for diagnosis in healthcare processes. Culcuoglu et al. (2012) propose a modified Kaizen approach that utilizes a series of two to four hour Kaizen Sessions for healthcare delivery systems. The authors also present how to document and measure the success of the sessions' effect.

2.3.13.3 Information Technologies in Healthcare

Lubitz and Wickramasinghe (2006) propose an integration of information technologies into healthcare operations. The proposed network-centric healthcare operations support system provide information flow among all users of the system and enable relevant knowledge to be generated and exchanged among the users. Later, Von Lubitz et al. (2008) extend the network-centric approach and give a new definition of Worldwide Healthcare Information Grid which is a global system for the efficiently conducting of healthcare operation around the world. Kuan (2009) proposes an RFID integrated healthcare system. The proposed service management system can handle operations such as revenue management, expenditure, service timing and provide analytics for managerial decision. Machado et al. (2010) focus on Ambient Intelligence based monitoring techniques in healthcare environments. They address the different methodologies put into operation in the healthcare sector, supported by a putative architecture which is used in various healthcare institutions to support RFID monitoring systems. Wu et al. (2011) suggest a remote healthcare platform for people with chronic disease living at home or inpatients living in the hospital. The proposed approach is based on open service platform and aims to overcome integration problem and improve ease of use. Shie et al. (2011) focus on electronic health information system and information exchange and give insight into the architecture, potential benefits, and challenges. The authors also present the current state of the applications while providing insight on the future research by highlighting the difficulties and potentials. Simonen et al. (2012) define the information concerning factors that promote the use of effectiveness data in healthcare management. The results show that the use of effectiveness data in healthcare management can be limited due to research, managerial work, and the organization. Grandinetti and Pisacane (2012) analyze the web services for healthcare management, and use operations research approaches in order to increase the effectiveness of the web services. Lillrank (2012) presents a definition of integration and coordination in health service production. The author identifies purposes, contexts, and design rules for integrated healthcare by applying a design science methodology. Chen (2012) integrates geographic information system, radio frequency identification and grid computing technology to provide an information system which can monitor and detect infectious events. Bhagya Lakshmi and Rajaram (2012) examine the impact of innovative approaches and information technology applications on the acceptance of rural healthcare services. Data is collected from 465 rural health personnel and analyzed statistically.

Liu and Park (2013) define the challenges and requirements of an e-Healthcare interconnection infrastructure and provide a framework. In their design, the authors take into account dimensions such as; security, integrated service management, on-demand access to the network, quality-of-service, accounting. Shi et al. (2013) investigate electronic health record systems and health information exchange by presenting their architecture, benefits, challenges, and other related issues. Lucas et al. (2013a) examine the relationship between facility management and healthcare delivery. A case study is used to define different types of information needed to

perform maintenance tasks satisfactorily. Lucas et al. (2013b) propose a healthcare facility information management prototype, which allows facility managers to respond more efficiently and effectively to facility related events. Sobol and Prater (2011, 2013) analyze healthcare practices of two countries, the United States and Taiwan, in order to provide a formal benchmark of information technology usage. The authors compare the countries regarding adaptation of information systems and the efficiency created by the usage. Banerjee et al. (2013) propose an architecture of a cloud-based healthcare application intending to serve patients in emergency conditions. Under the emergency, the patient's medical history can be tracked by the cloud system and better decisions can be given. Healthcare management is an important interest area of supply chain management literature. Chen and Chang (2013) propose using RFID technology to produce a healthcare monitoring system which can trace every event the same times as the events happen. The applicability of the system is demonstrated in a pilot study. Ker et al. (2014) focus on the efficiency of logistic systems of medicals and evaluate the effectiveness of information systems used in pharmacies. Browne et al. (2015) focus on primary healthcare organizations. The authors summarize innovative intervention called EQUIP which designed to improve the capacity of primary healthcare clinics. The authors also provide information about real-world examples from four different clinics, which adopt the intervention.

2.3.13.4 Tactical and Strategic Decision-Making in Healthcare

Lavy and Shohet (2007a) develop a decision-making model which can integrate various parameters into a facility management tactical and strategic decision-making process. In the study, the authors provide the architecture and procedures of the proposed model.

2.3.13.5 Lean Management in Healthcare

Enache-Pommer and Horman (2008) integrate lean approach with green principles and apply them in children' hospitals. The results of the application of the proposed approach in three hospitals show that with the proposed approach the hospitals may become more efficient and healthful. Hicks et al. (2015) provide a case study on applying a lean approach to production, preparation, and process (3P) to design a new endoscopy unit in England. The result of the analysis shows that this 3P participative design method is an effective tool for meeting the requirements of multiple stakeholders. Healthcare management uses the principles of system analysis and design.

2.3.13.6 Environmental Management in Healthcare

Kagioglou and Tzortzopoulos (2010) examine the business flow of healthcare infrastructure sector from a built environment perspective. The authors give insight into the trends in healthcare, explain the alternative models of healthcare delivery; current building and investment programs; clarify the procurement process, and describe facilities management activities, give details about related financial models, risk evaluation on healthcare. De Fátima Castro et al. (2015) propose a benchmarking approach for healthcare buildings. The authors take Building Sustainability Assessment methods as a baseline and try to reduce the subjectivity given in the definitions. As a result, the authors define criteria for healthcare buildings in four main groups namely, consumption of resources, waste production, costs and environmental impacts.

2.3.13.7 Risk Analysis in Healthcare

Sørup and Jacobsen (2013) focus on employee absence in the healthcare sector. The authors initially define the main factors of employee absence using satisfaction scores, and then they use these findings to form a management framework, which provides information about risk factors associated with employee absence. Technology management is also an important aspect of healthcare management.

2.4 Conclusions

In this chapter, we reviewed the literature review on healthcare management. The review results can be classified as follows: the works on quantitative and qualitative techniques in healthcare management; the works on healthcare case studies; and the works on healthcare literature reviews. We later classified the techniques and approaches used in healthcare research and gave some related recent papers under each class.

The common aims of the HCM studies in the literature can be classified as the quality of health care, efficiency improvement, patient satisfaction and financial performance. Literature review reveals that quantitative techniques are more frequently used than qualitative techniques in HCM. Within quantitative techniques, the most commonly used ones are simulation, statistical decision making and data mining. The common feature of these techniques is that they are data driven.

Data analysis is a critical component of healthcare management works. Data analysis has gained acceleration in the recent years from data mining to big data. Especially, with the emergence of IOT (internet of things) technologies, new sources

of data related to healthcare management are included in these studies. For instance, IOT and Big Data technologies will provide improved patient care, flexible patient monitoring, and improved drug management. These emerged technologies will take an important place in the future of HCM research areas. Hence, we suggest data mining and big data analyses to be utilized in healthcare for further studies.

References

Adasme P, Andrade R, Leung J, Lisser A (2015) A comparative study of network-based approaches for routing in healthcare wireless body area networks, ICORES 2015 – 4th International Conference on Operations Research and Enterprise Systems, Proceedings, pp 125–132

Ahsan K, Shah H, Kingston P (2010) Healthcare modelling through enterprise architecture: a hospital case ITNG2010 – 7th International Conference on Information Technology: New Generations, art. no. 5501732, pp 460–465

Aktaş E, Ülengin F, Önsel Şahin S (2007) A decision support system to improve the efficiency of resource allocation in healthcare management. Socio Econ Plan Sci 41(2):130–146

Al-Balushi S, Sohal AS, Singh PJ, Hajri AA, Farsi YMA, Abri RA (2014) Readiness factors for lean implementation in healthcare settings – a literature review. J Health Organ Manag 28(2):135–153

Al-Shayea AM (2011) Measuring hospital's units efficiency: a data envelopment analysis approach. Int J Eng Technol 11(6):7–14

Ashrafi MY, Karami M, Safdari R, Nazeri A (2013) Selective overview on decision support systems: focus on healthcare. Life Sci J 10(4):920–928

Banerjee A, Agrawal P, Rajkumar R (2013) Design of a cloud based emergency healthcare service model. Int J Appl Eng Res 8:2261–2264

Baril C, Gascon V, Miller J, Côté N (2016) Use of a discrete-event simulation in a Kaizen event: a case study in healthcare. Eur J Oper Res 249(1):327–339

Bauernschmitt E, Conradie W (2010) Investigation into current supply chain practices at a private healthcare provider in South Africa, Proceedings of the European Conference on Knowledge Management, ECKM, pp 54–64

Beran TN, Violato C (2010) Structural equation modeling in medical research: a primer. BMC Res Notes 3:267

Bhagya Lakshmi K, Rajaram M (2012) Impact of information technology reliance and innovativeness on rural healthcare services: study of Dindigul District in Tamilnadu, India. Telemed e-Health 18(5):360–370

Bora C, Miragliotta G, Pala E, Perego A, Tumino A (2011) A 6-step methodology to evaluate the cost and benefits of RFID-based innovation in the healthcare, Proceedings – 2011 10th International Conference on Mobile Business, ICMB 2011, art. no. 6047089, pp 352–361

Bose R (2003) Knowledge management-enabled health care management systems: capabilities, infrastructure, and decision-support. Expert Syst Appl 24(1):59–71

Boston-Fleischhauer C (2008) Enhancing healthcare process design with human factors engineering and reliability science, Part 1: setting the context. J Nurs Adm 38(1):27–32

Brailsford S (2005) Overcoming the barriers to implementation of operations research simulation models in healthcare. Clin Invest Med 28(6):312–315

Brailsford SC (2008) System dynamics: what's in it for healthcare simulation modelers, Proceedings – Winter Simulation Conference, art. no. 4736227, pp 1478–1483

Brailsford SC, Desai SM, Viana J (2010) Towards the holy grail: combining system dynamics and discrete-event simulation in healthcare Proceedings, Winter Simulation Conference, art. no. 5678927, pp 2293–2303

Browne AJ, Varcoe C, Ford-Gilboe M, Wathen CN (2015) EQUIP healthcare: an overview of a multi-component intervention to enhance equity-oriented care in primary health care settings. Int J Equity Health 14 (1), art. no. 271

Bullock A, Morris ZS, Atwell C (2013) Exchanging knowledge through healthcare manager placements in research teams. Serv Ind J 33(13–14):1363–1380

Buyurgan N, Farrokhvar P (2015) Supply chain-related adverse events and patient safety in healthcare. Int J Healthc Inf Syst Inform 10(2):14–33

Cabrera E, Taboada M, Iglesias M, Epelde F, Luque E (2012) Simulation optimization for healthcare emergency departments. Procedia Comput Sci 9:1464–1473

Chandra C (2008) The case for healthcare supply chain management: insights from problem-solving approaches. Int J Procure Manag 1(3):261–279

Chandrasekaran A, Senot C, Boyer KK (2012) Process management impact on clinical and experiential quality: managing tensions between safe and patient-centered healthcare. Manuf Serv Oper Manag 14(4):548–566

Chen C-C, Chang Y-F (2013) Smart healthcare environment: design with RFID technology and performance evaluation. J Med Biol Eng 33(4):427–432

Chen YJ (2012) GIS, grid computing and RFID in healthcare information supply chain: a case for infectious disaster management. In: Cases on supply chain and distribution management: issues and principles. IGI Global, Hershey, pp 256–268

Cox LA (2006) Quantitative health risk analysis methods: modeling the human health impacts of antibiotics used in food animals. Springer, New York

Culcuoglu MU, Wang S, Powers C, Hillman M (2012) A new approach to kaizen events in healthcare delivery systems: Kaizen sessions, 62nd IIE Annual Conference and Expo 2012, pp 2699–2707

Danas K, Roudsari A, Ketikidis PH (2006) The applicability of a multi-attribute classification framework in the healthcare industry. J Manuf Technol Manag 17(6):772–785

Darrel A, Hume M, Hardie T, Soar J (2014) The benefits of big data analytics in the healthcare sector: what are they and who benefits? Big data analytics in bioinformatics and healthcare, pp 406–439

Davey S, Raghav SK, Singh JV, Davey A, Singh N (2015) A comparative evaluation of public health centers with private health training centers on primary healthcare parameters in India: a study by data envelopment analysis technique. Indian J Community Med 40(4):252–257

De Fátima Castro M, Mateus R, Serôdio F, Bragança L (2015) Development of benchmarks for operating costs and resources consumption to be used in healthcare building sustainability assessment methods. Sustainability (Switzerland) 7(10):13222–13248

De Mast J, Kemper B, Does RJMM, Mandjes M, Van Der Bijl Y (2011) Process improvement in healthcare: overall resource efficiency. Qual Reliab Eng Int 27(8):1095–1106

Dehe B, Bamford D (2015) Development, test and comparison of two Multiple Criteria Decision Analysis (MCDA) models: a case of healthcare infrastructure location. Expert Syst Appl 42(19):6717–6727

Devaraj S, Ow TT, Kohli R (2013) Examining the impact of information technology and patient flow on healthcare performance: a Theory of Swift and Even Flow (TSEF) perspective. J Oper Manag 31(4):181–192

Dey PK, Hariharan S (2006) Integrated approach to healthcare quality management: a case study. TQM Mag 18(6):583–605

Dobrzykowski D, Saboori Deilami V, Hong P, Kim S-C (2014) A structured analysis of operations and supply chain management research in healthcare (1982–2011). Int J Prod Econ 147(PART B):514–530

Dobrzykowski DD, Tarafdar M (2015) Understanding information exchange in healthcare operations: evidence from hospitals and patients. J Oper Manag 36:201–214

Dortland MR, Dewulf G, Voordijk H (2013) Project coalitions in healthcare construction projects and the application of real options: an exploratory survey. Health Environ Res Des J 7(1):14–36

Earnshaw SR, Dennett SL (2003) Integer/linear mathematical programming models: a tool for allocating healthcare resources. PharmacoEconomics 21(12):839–851

Eldabi T, Paul RJ, Young T (2007) Simulation modelling in healthcare: reviewing legacies and investigating futures. J Oper Res Soc 58(2):262–270

Enache-Pommer E, Horman M (2008) Greening of healthcare facilities: case studies of children's hospitals, Proceedings of the AEI 2008 Conference – AEI 2008: Building Integration Solutions, 328.

Fakhimi M, Mustafee N (2012) Application of operations research within the UK healthcare context, Operational Research Society Simulation Workshop, SW 2012, pp 66–82

Fakhimi M, Probert J (2013) Operations research within UK healthcare: a review. J Enterp Inf Manag 26(1):21–49

Farinella D, Saitta P, Signorino GM (2011) The regionalization of the public health system and new governance models for healthcare: the stroke network case in Italy World Medical and Health Policy 3(2), art. no. 6

Fervers L, Oser P, Picot G (2015) Globalization and healthcare policy: a constraint on growing expenditures. J Eur Publ Policy 20 p

Finstuen K, Mangelsdorff AD (2006) Executive competencies in healthcare administration: preceptors of the Army-Baylor University Graduate Program. J Health Adm Educ 23(2): 199–215

Fortsch SM, Khapalova EA (2016) Reducing uncertainty in demand for blood. Oper Res Health Care 9:16–28

Foster EM, Hosking MR, Ziya S (2010) A spoonful of math helps the medicine go down: an illustration of how healthcare can benefit from mathematical modeling and analysis. BMC Med Res Methodol 10, art. no. 60

Gaion S, Fanti MP, Mininel S, Ukovich W, Vatta F (2009) Modelling and validation of alarm management workflow in healthcare integrating IHE-PCD profile and Coloured Petri Nets IFAC Proceedings Volumes (IFAC-PapersOnline), 2(PART 1), pp 169–174

Garg L, McClean SI, Barton M, Meenan BJ, Fullerton K (2012) Intelligent patient management and resource planning for complex, heterogeneous, and stochastic healthcare systems. IEEE Trans Syst Man Cybern A: Syst Hum 42(6):1332–1345

Gonsalves T, Itoh K (2009) Service optimization with patient satisfaction in healthcare systems. J Simul 3(3):150–162

Gowen CR III, Mcfadden KL, Hoobler JM, Tallon WJ (2006) Exploring the efficacy of healthcare quality practices, employee commitment, and employee control. J Oper Manag 24(6):765–778

Grandinetti L, Pisacane O (2012) Web services for healthcare management. In: Technologies and protocols for the future of internet design: reinventing the web. Information Science Reference, Hershey, pp 60–94

Green LV (2012) The vital role of operations analysis in ımproving healthcare delivery. Manuf Serv Oper Manag 14(4):488–494

Guo SJ, Hsu CH, Lin CJ (2008) A study of community healthcare competency among public health nurses. J Nurs Res JNR 16(4):286–296

Hadziabdic E, Heikkilä K, Albin B, Hjelm K (2011) Problems and consequences in the use of professional interpreters: qualitative analysis of incidents from primary healthcare. Nurs Inq 18(3):253–261

Haggerty JL, Lévesque J-F, Santor DA, Burge F, Beaulieu C, Bouharaoui F, Beaulieu M-D, Pineault R, Gass D (2011) Accessibility from the patient perspective: comparison of primary healthcare evaluation instruments. Healthcare Policy 7(SPEC. ISSUE):94–107

Hicks C, McGovern T, Prior G, Smith I (2015) Applying lean principles to the design of healthcare facilities. Int J Prod Econ 170:677–686

Homer JB, Hirsch GB (2006) System dynamics modeling for public health: background and opportunities. Am J Public Health 96(3):452–458

Huang J, Jennings NR, Fox J (1995) Agent-based approach to health care management. Appl Artif Intell 9(4):401–420

Hunt MR (2009) Moral experience of Canadian healthcare professionals in humanitarian work. Prehosp Disaster Med 24(6):518–524

Hussain M, Malik M, Al Neyadi HS (2016) AHP framework to assist lean deployment in Abu Dhabi public healthcare delivery system. Bus Process Manag J 22(3):546–565

Jaafaripooyan E, Agrizzi D, Akbari-Haghighi F (2011) Healthcare accreditation systems: further perspectives on performance measures. Int J Qual Health Care 23(6.), art. no. mzr063):645–656

Johnson J (2008) Simple queuing theory tools you can use in healthcare Society for Health Systems Conference and Expo, 62 p

Kagioglou M, Tzortzopoulos P (2010) Improving healthcare through built environment infrastructure, improving healthcare through built environment infrastructure, 272 p

Kamarudeen S (2010) Amenable mortality as an indicator of healthcare quality – a literature review. Health Stat Q/Off Natl Stat 47:66–80

Ker J-I, Wang Y, Hajli MN, Song J, Ker CW (2014) Deploying lean in healthcare: evaluating information technology effectiveness in U.S. hospital pharmacies. Int J Inf Manag 34(4):556–560

Knight VA, Williams JE, Reynolds I (2012) Modelling patient choice in healthcare systems: development and application of a discrete event simulation with agent-based decision making. J Simul 6(2):92–102

Kontio E, Kontio J, Korvenranta H, Salanterä S (2007) How do the healthcare information systems (HIS) help: a literature review. ECIME 2007: European Conference on Information Management and Evaluation, pp 287–294

Krey M, Harriehausen B, Knoll M, Furnell S (2010) IT governance and its impact on the Swiss healthcare, UKSim2010 – UKSim 12th International Conference on Computer Modelling and Simulation, art. no. 5480473

Kuan JH (2009) Development of healthcare service management system using systems engineering and RFID technology, 19th Annual International Symposium of the International Council on Systems Engineering, INCOSE 2009, 2, pp 828–843

Landa P, Aringhieri R, Soriano P, Tànfani E, Testi A (2016) A hybrid optimization algorithm for surgeries scheduling. Oper Res Health Care 8:103–114

Lane DC, Monefeldt C, Rosenhead JV (2000) Looking in the wrong place for healthcare improvements: a system dynamics study of an accident and emergency department. J Oper Res Soc 51(5):518–531

Lavis J, Davies HTO, Oxman A, Denis J-L, Golden-Biddle K, Ferlie E (2005) Towards systematic reviews that inform health care management and policy-making. J Health Serv Res Policy 10:35–48

Lavy S, Shohet IM (2007a) A strategic integrated healthcare facility management model. Int J Strateg Prop Manag 11(3):125–142

Lavy S, Shohet IM (2007b) On the effect of service life conditions on the maintenance costs of healthcare facilities. Constr Manag Econ 25(10):1087–1098

Lavy S, Shohet IM (2009) Integrated healthcare facilities maintenance management model: case studies. Facilities 27(3–4):107–119

Lillrank P (2012) Integration and coordination in healthcare: an operations management view. (2012. J Integr Care 20(1):6–12

Liu W, Park EK (2013) e-Healthcare interconnection networking services. J Commun 8(9):550–560

Lubitz DV, Wickramasinghe N (2006) Networkcentric healthcare and bioinformatics: unified operations within three domains of knowledge. Expert Syst Appl 30(1):11–23

Lucas J, Bulbul T, Thabet W (2013a) A pilot model for a proof of concept healthcare facility information management prototype. J Inf Technol Constr 18:76–98

Lucas J, Bulbul T, Thabet W, Anumba C (2013b) Case analysis to identify information links between facility management and healthcare delivery information in a hospital setting. J Archit Eng 19(2):134–145

Machado J, Miranda M, Gonçalves P, Abelha A, Neves J, Marques A (2010) AIDATrace – interoperation platform for active monitoring in healthcare environments 8th International Industrial Simulation Conference 2010, ISC 2010, pp 67–71

Malhotra R, Lehrman S, Malhotra DK (2015) Benchmarking managed healthcare companies. Appl Manag Sci 17:203–222

Marcarelli G (2016) Multi-criteria decision making for evaluating healthcare policies: the benefit/cost analysis by the analytic hierarchy process. Int J Med Eng Inform 8(2):163–181

Matopoulos A, Michailidou L (2013) Implementing collaborative practices in the healthcare supply chain: insights into hospital-vendor operations. Int J Logist Syst Manag 15(2–3):288–303

McAlearney AS, Garman AN, Song PH, McHugh M, Robbins J, Harrison MI (2010) Supporting those who dare to care: five case studies of high-performance work practices in healthcare, Academy of Management 2010 Annual Meeting – Dare to Care: Passion and Compassion in Management Practice and Research, AOM 2010, 6 p

Michailidou M, Goldstein SD, Sacco Casamassima MG, Salazar JH, Elliott R, Hundt J, Abdullah F (2015) Laparoscopic versus open appendectomy in children: the effect of surgical technique on healthcare costs. Am J Surg 210(2):270–275

Mohammadi SM, Mohammadi SF, Hedges JR (2007) Conceptualizing a quality plan for healthcare: a philosophical reflection on the relevance of the health profession to society. Health Care Anal 15(4):337–361

Motamarri S, Akter S, Ray P, Tseng C-L (2014) Distinguishing "mHealth" from other healthcare services in a developing country: a study from the service quality perspective. Commun Assoc Inf Syst 34(1.), art. no. 34):669–692

Nelson BA (2011) Healthcare team members' perception of staffing adequacy in a comprehensive cancer center. Oncol Nurs Forum 38(1):52–59

Nemeth C, Cook R (2010) Infusing healthcare with resilience, 20th Annual International Symposium of the International Council on Systems Engineering, INCOSE 2010, 2, pp 1073–1087

Nicholson L, Vakharia AJ, Selcuk Erenguc S (2004) Outsourcing inventory management decisions in healthcare: models and application. Eur J Oper Res 154(1):271–290

Ozcan YA (2009) Quantitative methods in health care management: techniques and applications. Wiley

Papadopoulos T, Radnor Z, Merali Y (2011) The role of actor associations in understanding the implementation of Lean thinking in healthcare. Int J Oper Prod Manag 31(2):167–191

Paschou M, Papadimitiriou C, Nodarakis N, Korezelidis K, Sakkopoulos E, Tsakalidis A (2015) Enhanced healthcare personnel rostering solution using mobile technologies. J Syst Softw 100:44–53

Rais A, Vianaa A (2011) Operations research in healthcare: a survey. Int Trans Oper Res 18(1): 1–31

Ramakrishnan S, Courtney AK, Srihari K, Emick F (2005) Streamlining the workflow at the CT scan area of a healthcare provider. IIE Annual Conference and Exposition 2005, 6 p

Ramakrishnan S, Nagarkar K, DeGennaro M, Srihari K, Courtney AK, Emick F (2004) A study of the CT Scan area of a healthcare provider. Proceedings – Winter Simulation Conference, 2, pp 2025–2031

Ramírez-Ríos DG, Manotas Romero LP, Paez-Logreira H, Ramírez L, Florez YAJ (2015) Data mining for the unique identification of patients in the national healthcare systems. ICORES 2015 – 4th International Conference on Operations Research and Enterprise Systems, Proceedings, pp 211–217

Rich N, Piercy N (2013) Losing patients: a systems view on healthcare improvement. Prod Plan Control 24(10–11):962–975

Robert G, Stevens A, Gabbay J (1999) 'Early warning systems' for identifying new healthcare Technologies. Health Technol Assess 3(13):iii–i99

Robinson S, Radnor ZJ, Burgess N, Worthington C (2012) SimLean: utilising simulation in the implementation of lean in healthcare. Eur J Oper Res 219(1):188–197

Roshanaei V, Luong C, Aleman DM, Urbach D (2017) Propagating logic-based Benders' decomposition approaches for distributed operating room scheduling. Eur J Oper Res 257(2):439–455

Rönnberg E, Larsson T (2010) Automating the self-scheduling process of nurses in Swedish healthcare: a pilot study. Health Care Manag Sci 13(1):35–53

Sachdeva R, Williams T, Quigley J (2007) Mixing methodologies to enhance the implementation of healthcare operational research. J Oper Res Soc 58(2):159–167
Sadatsafavi H, Walewski J, Shepley MM (2015) Factors influencing evaluation of patient areas, work spaces, and staff areas by healthcare professionals. Indoor Built Environ 24(4):439–456
Saedi S, Erhun Kundakcioglu O, Henry AC (2016) Mitigating the impact of drug shortages for a healthcare facility: an inventory management approach. Eur J Oper Res 251(1):107–123
Samuel C, Gonapa K, Chaudhary PK, Mishra A (2010) Supply chain dynamics in healthcare services. Int J Health Care Qual Assur 23(7):631–642
Savsar M, Al-Ajmi MM (2012) A quality control application in healthcare management using experimental design criteria. Int J Health Care Qual Assur 25(1):53–63
Shi J, Upadhyaya S, Erdem E (2013) Health information exchange for improving the efficiency and quality of healthcare delivery. User-Driven Health Concepts Methodol Tools Appl 2:714–736
Shi J, Upadhyaya S, Erdem E (2011) Health information exchange for improving the efficiency and quality of healthcare delivery. In: Management engineering for effective healthcare delivery: principles and applications. IGI Global, Hershey, pp 352–374
Simonen O, Viitanen E, Blom M (2012) Factors relating to effectiveness data use in healthcare management. Int J Product Perform Manag 61(7):752–764
Sobol M, Prater E (2011) Adoption, usage and efficiency: benchmarking healthcare IT in private practices. Int J Healthc Inf Syst Inform 6(1):36–50
Sobol M, Prater E (2013) Adoption, usage and efficiency: benchmarking healthcare it in private practices. In: Healthcare information technology innovation and sustainability: frontiers and adoption. Medical Information Science Reference, Hershey, pp 145–159
Sørup CM, Jacobsen P (2013) Healthcare performance turned into decision support. J Health Organ Manag 27(1):64–84
Swangnetr M, Zhu B, Kaber D, Taylor K (2010) Meta-analysis of user age and service robot configuration effects on human-robot interaction in a healthcare application, AAAI Fall Symposium – Technical Report, FS-10-05, pp 121–126
Tanfani E, Testi A (2012) Advanced decision making methods applied to health care. Springer-Verlag, Italia
Uddin J, Koehlmoos TP, Saha NC, Islam Z Khan IA, Quaiyum MA (2012) Strategies for providing healthcare services to street-dwellers in Dhaka city: evidence from an operations research. Health Res Policy Syst 10, art. no. 19
Utter GH, Cuny J, Sama P, Silver MR, Zrelak PA, Baron R, Drsler SE, Romano PS (2010) Detection of postoperative respiratory failure: how predictive is the agency for healthcare research and quality's patient safety indicator? J Am Coll Surg 211(3):347–354.e29
Van De Weerdt C, Baratta R (2012) New working conditions and consequences on activity of home healthcare workers. Work 41(SUPPL.1):1–4
Van Minh H, Anh TT, Rocklöv J, Giang KB, Trang LQ, Sahlen K-G, Nilsson M, Weinehall L (2014) Primary healthcare system capacities for responding to storm and flood-related health problems: a case study from a rural district in central Vietnam. Global Health Action 7(1), art. no. 23007
van Vactor JD (2011) A case study of collaborative communications within healthcare logistics. Leadersh Health Serv 24(1):51–63
Von Lubitz D, Patricelli F, Palma P (2008) Telecommunications infrastructure for worldwide networkcentric healthcare operations and the associated information system. Int J Bus Syst Res 2(1):34–51
Vozikis A, Goulionis JE, Benos VK (2012) The partially observable Markov decision processes in healthcare: an application to patients with ischemic heart disease (IHD). Oper Res 12(1):3–14
Wang J, Kamin C, Glover WJ (2011) A survey of systems and improvement approaches for psychological healthcare. 61st Annual IIE Conference and Expo Proceedings
Weidmer BA, Cleary PD, Keller S, Evensen C, Hurtado MP, Kosiak B, Gallagher PM, Levine R, Hays RD (2014) Development and evaluation of the CAHPS (Consumer Assessment of Healthcare Providers and Systems) survey for in-center hemodialysis patients. Am J Kidney Dis 64(5):753–760

Wu JN, Chen PY, Wu NC, Chao YC (2011) Event-driven open service platform based remote healthcare system International Conference on Advanced Communication Technology, ICACT, pp 1559–1562

Xie X, Lawley MA (2015) Operations research in healthcare. Int J Prod Res 53(24):7173–7176

Zadeh R, Sadatsafavi H, Xue R (2015) Evidence-based and value-based decision making about healthcare design: an economic evaluation of the safety and quality outcomes. Health Environ Res Des J 8(4):58–76

Zamora Aguas JP, Adarme WA, Arango Serna MD (2013) Supply risk analysis: applying system dynamics to the Colombian healthcare sector [Análisis del riesgo de aprovisionamiento: Caso aplicado al sector salud Colombiano, utilizando la dinámica de sistemas]. Ing Investig 33(3):76–81

Zeng Q-L, Li D-D, Yang Y-B (2013) VIKOR method with enhanced accuracy for multiple criteria decision making in healthcare management. J Med Syst 37(2), art. no. 9908

Zonca P, Moravcova V, Maresova P, Kuca K (2015) Competition in healthcare in Czech Republic with focus on surgery. Proceedings of the 25th International Business Information Management Association Conference – Innovation Vision 2020: from Regional Development Sustainability to Global Economic Growth, IBIMA 2015, pp 846–854

Part II
Medical Units in Hospital

Chapter 3
The Real Time Management of Operating Rooms

Davide Duma and Roberto Aringhieri

3.1 Introduction

Operating Room (OR) planning and scheduling is a research topic widely discussed in the literature. Cardoen et al. (2010) and Guerriero and Guido (2011) provide an exhaustive review of the problems belonging to this topic, also analyzing in detail multiple fields related to the problem settings and summarizing significant trends in research areas of future interest. Such problems are usually classified into three phases corresponding to three decision levels (Testi et al. 2007), namely strategic (long term), tactical (medium term) and operational (short term).

At the operational decision level, the problem arising in the OR planning is also called "surgery process scheduling", which usually consists in (1) selecting elective patients from an usually long waiting list and assigning them to a specific OR time session (i.e., an operating room on a specific day) over a planning horizon, (2) determining the precise sequence of surgical procedures and the allocation of resources for each OR time session, and (3) dealing with the arrival of non-elective patients requiring a surgery within a given time threshold.

The Real Time Management (RTM) of operating rooms is the decision problem arising during the fulfilment of the surgery process scheduling of elective and non-elective patients, that is the problem of supervising the execution of such a schedule and, in case of delays, to take the more rational decision regarding the surgery cancellation or the overtime assignment. The RTM is characterized by the uncertainty of its main parameters, that is, the duration of a surgery and the arrival of non-elective patients. The RTM could deal with different objectives,

D. Duma • R. Aringhieri (✉)
Dipartimento di Informatica, Università degli Studi di Torino, Corso Svizzera 185, I-10149 Torino, Italy
e-mail: davide.duma@unito.it; roberto.aringhieri@unito.it

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_3

that is to maximize the operating room utilization and to minimize the number of surgeries cancelled. Especially when considering the inherent stochasticity of the problem (Bruni et al. 2015; Addis et al. 2016; Landa et al. 2016; van Veen-Berkx et al. 2016), the two objectives are conflicting (Beaulieu et al. 2012).

The optimization literature reports few attempts to address the problem as reported in Hans and Vanberkel (2012). The problem of rescheduling the elective patients upon the arrival of emergency patients is addressed in Erdem et al. (2011, 2012). The authors proposed a MILP model which considers the overtime cost of the operating rooms and/or the post-anesthesia care units, the cost of postponing or preponing elective surgeries, and the cost of turning down the emergency patients. They proposed a genetic algorithm for its approximate and faster solution. The results of the case study suggest that, instead of shuffling the elective surgeries, it would be worthwhile to consider performing the elective surgeries using the overtime of the operating rooms. Note that the problem of rescheduling patients can be addressed as a particular job shop scheduling problem (see, e.g., Pham and Klinkert 2008; Stuart and Kozan 2012) but these experiences can not be directly applied to the operating room context due to its peculiarity in the evaluation of a solution, as we will show along the chapter.

The current development of the health care systems is aimed to recognize the central role of the patient as opposed to the one of the health care providers. In this context, Clinical Pathways (CPs) shift the attention from a single health benefit to the health care chain that starts to resolve the illness episode. They can be defined as "health-care structured multidisciplinary plans that describe spatial and temporal sequences of activities to be performed, based on the scientific and technical knowledge and the organizational, professional and technological available resources" (Campbell et al. 1998). A comprehensive approach to the surgery process scheduling of only elective patients has been discussed in Duma and Aringhieri (2015), in which the authors proposed also an online algorithm for the RTM. In that paper, the authors provided an extensive analysis of the impact of several optimization procedures on the performance of a surgical CP estimated exploiting a series of patient- and facility-centered indices as previously discussed in Aringhieri and Duma (2016).

Starting from the problem with only elective patients, it could be of great interest, both from a methodological and managerial point of view, to consider in our analysis two crucial aspects of the OR management, that is the management of non-elective patients and the resource sharing among different surgical pathways. These two aspects make the starting problem more challenging requiring the adoption of unconventional solution methodologies (Aringhieri et al. 2013).

While the list of the elective patients is known at the moment of the surgery process scheduling, the arrivals of the non-elective patients are unknown. Non-elective patients differs from elective ones by their request of a compulsory surgery within a tight time limit, usually ranging from "*as soon as possible*" to "*within 24 h*". Thus the management of non-elective patients poses a challenge problem since delaying their surgery may increase the risk of postoperative complications and morbidity. As reported in literature (see, e.g., Van Riet and Demeulemeester

2015), two main policies can be adopted to deal with non-elective patients, that is to have a certain number of dedicated ORs (see, e.g., Heng and Wright 2013) or to share available ORs (see, e.g., Wullink et al. 2007). From the RTM point of view, the sharing OR policy is more interesting to study because of the need to establish how to insert non-elective patients within OR sessions on which the elective patients have been already scheduled. Therefore, the need of making available an OR session within a tight time limit and without an ex-ante planning improves the complexity of supervising the execution of the surgery schedule affecting both the number of cancellations of elective surgery and the use of the overtime.

Among different surgical pathways, the critical resources that can be shared in order to improve the pathway management, are the overtime and the OR sessions. The master surgical schedule usually defines the specific assignment of OR sessions to different surgical pathways (Testi and Tànfani 2009), and it should be updated whenever the total amount of the OR time or the requirements of some surgical pathways change. This can occur not only as a response to long term changes in the overall OR capacity or staffing fluctuations, but also in response to seasonal fluctuations in demand (see, e.g., Banditori et al. 2013; van Oostrum et al. 2008). The objective of the resource sharing is therefore to have a fair assignment of both the overtime and the OR sessions to different surgical pathways.

In this chapter we propose online optimization approaches for the RTM capable to deal with (1) the elective and non-elective patient flows within a single surgical pathway, and with (2) the resource sharing among different surgical pathways of elective patients. The chapter is organized as follows.

Section 3.2 describes the problem of dealing with the elective and non-elective patient flows in a single surgical pathway. The online approach for the RTM for a joint flow of elective and non-elective patients is discussed in Sect. 3.3. In Sect. 3.4, we also provide a mixed-integer programming model to compute the optimal offline solution, that is the optimal solution assuming to know in advance all the information that are acquired over time by the online solution. Such a solution provides a significant contribution to evaluate the competitiveness of the online approach. The quantitative analysis to prove the effectiveness of the proposed approach is reported in Sect. 3.5.

Section 3.6 introduces the problem of sharing resources among different surgical pathways, highlighting the differences with respect to the problem discussed in Sect. 3.3. Different and alternative policies for the management of the ORs and the overtime are proposed in Sects. 3.6.1 and 3.6.2, respectively. Following the same framework of analysis introduced in Sect. 3.5.1, we determine the best policy combination to manage the shared resources in Sect. 3.6.3.

General remarks and conclusions are discussed in Sect. 3.7.

3.2 The Management of Elective and Non-elective Patients

In this section we would describe the operative context that we are considering in order to better motivate our approach. An ex-ante scheduling is usually performed before the starting of the planning time horizon, which is usually set to 1 week in the literature. Here we consider the following version of the surgery process scheduling: for each OR session available, it determines which elective patients should be operated on (*advanced scheduling*) selecting them from the pre-admission list, and the sequence of their surgical interventions (*allocation scheduling*). We consider the surgery process scheduling of a single specialty under the block scheduling or closed block planning paradigm (van Oostrum et al. 2010): for each planning period, a number of OR time blocks are assigned to the specialty, which schedules their surgical cases within these time blocks.

Let us consider the set J of the operating rooms and let K be the set of the days of the week. Then, let $S \subseteq J \times K$ be the set of the OR sessions (j, k), each one denoting an OR $j \in J$ available the day $k \in K$ and having duration equal to d_{jk} minutes. The overall number of OR sessions is $n = |S|$. Let I be the set of the elective patients in the pre-admission list and $L \subseteq I$ the set of scheduled patients to be operated on during the next planning time horizon. Note that L can be partitioned in n subsets $L_{jk} \subseteq L$ where L_{jk} is the set of the patient scheduled on the OR session (j, k).

For each patient $i \in I$ the Estimated Operating Time (EOT) e_i and the number of days elapsed in the waiting list t_i are known. After the surgery, also the Real Operating Time (ROT) r_i will be known. Another relevant information is the Diagnosis Related Group (DRG). A DRG defines a general time limit expressed in days before which the patient should be operated on (i.e., days to surgery). In our context, a *Urgency Related Group* (URG) is assigned to each patient belonging to the same DRG: the URG states a more accurate time limit called *Maximum Time Before Treatment* (MTBT) and denoted by $t_i^{\max}$. In other words, URG allows to define a partition of the patients in the same DRG in order to prioritize their surgical operation.

Usually, the advanced scheduling has the objective to maximize the utilization of each OR session considering its overall duration d_{jk} promoting the selection from I of those patients whose t_i is closer to $t_i^{\max}$. Since the ROT r_i is unknown, such a scheduling is usually performed taking into account the EOT e_i determined by the physician during the outpatient visit. Note that the EOTs are also used in the allocation scheduling to determine the patient sequence within each OR session.

The RTM arises during the accomplishment of the surgery process schedule. Considering the OR session (j, k), it could happen that $r_i > e_i$ for the patient $i \in L_{jk}$. When this occurs, the schedule could be delayed and the overall delay could determine the exceeding of the OR session duration d_{jk}. In this case, two possible decisions can be considered, that is that of postponing a surgery or that of assigning a part of the available overtime Ω to allow the completion of the surgery plan. Such decisions have to take into account the need of guarantee the patient surgery before his/her MTBT but, at the same time, to avoid an over-allocation determining

Table 3.1 Notation

J: set of operating rooms	K: set of the days of the week
j: index of the operating room	k: index of the day
S: set of OR sessions	n: number of OR sessions
d_{jk}: duration of OR session (j,k)	Ω: weekly overtime available
I: set of patients in the pre-admission waiting list	L: set of scheduled patients
L_{jk}: set of patients scheduled into the OR session (j,k)	$t_i^{\max}$: MTBT of patient i
t_i: waiting days to surgery of the i-th patient	e_i: EOT of patient i
r_i: ROT of patient i	

a possible overtime failure in the next planning days. The RTM requires an online approach because the overtime demand until the end of the planning horizon is unknown. Further, the RTM is challenged by the unforeseeable arrivals of non-elective patients that must be operated on within their tight time limit. Unlike the elective ones, the non-elective patients are characterized only by the values e_i and r_i. Table 3.1 summarizes the notation introduced in this section.

3.3 *Ex-Ante* Approach: The Online Solution

In this section we introduce our online approach to deal with elective and non-elective patients.

3.3.1 The RTM with only Elective Patients

Let us suppose to consider the generic OR session $(j,k) \in S$. Let ρ_{jk}^{τ} be the time elapsed in the OR session (j,k) from the beginning of the session at the time τ. If the surgery of the m-th patient belonging to the schedule of L_{jk} ends at time τ, the effective time elapsed to operate on the first m patients is

$$\rho_{jk}^{\tau} = \sum_{i=i_1,\ldots,i_m} r_i. \tag{3.1}$$

Let us introduce the following parameter:

$$\beta_k^{\tau} = 1 + \frac{n_k}{n} - \frac{\Omega_k^{\tau}}{\Omega} \tag{3.2}$$

where Ω_k^{τ} is the remaining overtime at the time τ and n_k is the number of OR sessions from the day after k, that is

$$n_k = \left|\left\{(j',k') \in S : k' > k\right\}\right|.$$

The value β_k^τ would measure the overtime still available with respect to the number of OR sessions to be still performed. Actually, β_k^τ is close to 1 when the overtime has been used proportionally; it is between 0 and 1 or it is greater than 1 when it is underused or overused, respectively. Because of n_k is equal to 0 when k is the last day of the planning, we remark that β_k^τ is always less than or equal to 1 hence promoting the use of overtime.

In order to establish when the cancellation of a patient could lead to the exceeding of the MTBT, we define the parameter

$$\widetilde{w}_i = \frac{t_i + \delta}{t_i^{\max}} \tag{3.3}$$

where δ is the number of days until the beginning of the next planning horizon. It is worth noting that we would to avoid the cancellation of patients with $\widetilde{w}_i > 1$ because the rescheduling of such patients in the next planning horizon would exceed the MTBT.

Let $L_{jk}^\tau \subset L_{jk}$ be the set composed of the patients in L_{jk} still waiting for a surgery at the time τ, that is $L_{jk} \smallsetminus \{i_1, \ldots, i_m\}$. The online algorithm for the RTM with only elective patients starts every time a surgery ends and

$$\rho_{jk}^\tau + \sum_{i \in L_{jk}^\tau} e_i > d_{jk}. \tag{3.4}$$

It consists of two procedures.

Resequencing. The objective of the resequencing is twofold, that is to ensure the surgery of those patients close to their MTBT, and to maximize the OR utilization. To this end, patients are ordered in such a way to put first the patients having $\widetilde{w}_i > 1$; then, if such patients does not run out the available operating time d_{jk}, a subset of the remaining patients are selected to maximize the OR utilization trough a dynamic programming approach, which is an adaptation of that discussed in Sect. 3.4.1 of Martello and Toth (1990); in any case, the unselected patients are inserted at the end of the schedule since they could be operated on using the overtime;

Overtime allocation. Let i_{m+1} be the next patient in the schedule. Then, if

$$\rho_{jk}^\tau + e_{i_{m+1}} > d_{jk}\,,$$

the patient i_{m+1} could incur in a cancellation. Therefore, the required overtime is assigned to patient i_{m+1} if the overtime available is sufficient and at least one of the two following conditions is satisfied:

$$\widetilde{w}_i > 1\,, \tag{3.5}$$

$$\beta_k^\tau \left(\frac{e_{i_{m+1}} + \rho_{jk}^\tau}{d_{jk}} \right) \leq 1\,. \tag{3.6}$$

Otherwise the surgery of the patient i_{m+1} is postponed to the next week. This assumption is justified by the high level of utilization imposed by the advanced scheduling, which is reported in Sect. 3.5.1.

3.3.2 The RTM with Elective and Non-elective Patients

Non-elective patients should require to be operated on within different but usually tight time limits depending on their urgency. Such time limits can range from "*as soon as possible*" to "*within 24 h*" (see Table 4 in Van Riet and Demeulemeester 2015). When the time limit is very short, that is few hours, the most reasonable decision is to schedule the intervention as soon as possible. This situation can be handled by the algorithm for the RTM reported in Sect. 3.3.1 by properly increasing the available overtime (Duma and Aringhieri 2015).

Here we consider the arrivals of non-elective patients with a time limit of 24 h, that is they must be treated within the end of the current day k. In this context, the online approach has to decide in which OR session the non-elective patient has to be scheduled. Such a decision can determine a different need of overtime or the cancellation of the elective patients previously scheduled.

Let $S_k \subseteq S$ be the set of the OR sessions planned on the day k. Let us also to introduce the set Q^τ composed of the non-elective patients still waiting for a surgery at the time τ, respectively. Therefore the next surgery should be selected from the set $L^\tau_{jk} \cup Q^\tau$.

To deal with this online decision we develop the Non-Elective Worst Fit (NEWF) algorithm, which is a greedy construction of an alternative schedule of the patients in which we try to insert the non-elective patients in Q^τ, only when $Q^\tau \neq \emptyset$. On the basis of this alternative schedule, the online algorithm NEWF establishes the next patient to be operated on during the OR session (j, k), that is to continue with the planned schedule or to insert a non-elective patient in (j, k).

The NEWF operates in the following online context: on the day k, we consider the instant during which the surgery of the patient $i_m \in L_{hk}$ ends; at that moment, let μ_{jk} be the estimated time remaining to the end of the ongoing surgery in the OR session (j, k) (note that $\mu_{hk} = 0$). An example of the context in which the NEWF operates is reported in Fig. 3.1.

The pseudocode reported in Algorithm 1 describes the algorithm NEWF. After the initialization of the auxiliary data structures, the algorithm starts a loop to determine the alternative schedule. At each iteration, the current non-elective patient p^{ne} is scheduled on the OR session (j, k) which minimizes the difference between the estimated total duration of the operated and non-operated patients in L_{jk}

$$\rho^\tau_{jk} + \mu_{jk} + \sum_{i \in L'_j} e_i \tag{3.7}$$

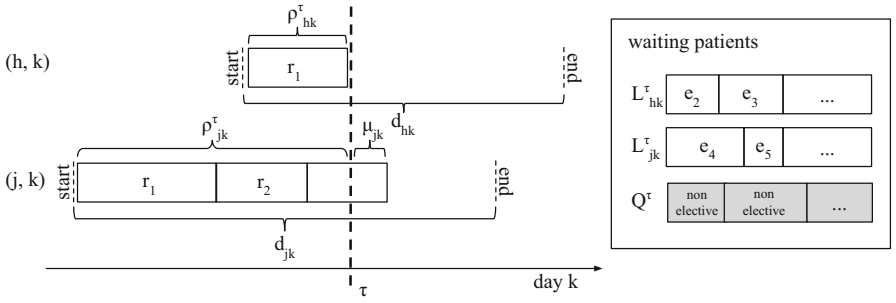

Fig. 3.1 Example of the context in which the NEWF operates

Algorithm 1 Non-elective Worst Fit

procedure NON-ELECTIVE WORST FIT
initialization:
  $p^e \leftarrow$ nextpatientin L^τ_{hk}; $Q' \leftarrow Q^m$;
  for each OR session (j,k) **do** $L'_j \leftarrow L^\tau_{jk}$;
  end for
loop:
  $p^{ne} \leftarrow$ patientin Q' withthemaximumwaitingtime;
  $j^\star \leftarrow \arg\min_{(j,k)} \left(\rho^\tau_{jk} + \mu_{jk} + \sum_{i \in L'_j} e_i - d_{jk}\right)$;
  if $j^\star = h$ **then return** p^{ne};
  end if
  $L'_{j^\star} \leftarrow L'_{j^\star} \cup \{p^{ne}\}$; $Q' \leftarrow Q' \setminus \{p^{ne}\}$;
  if $Q' \neq \emptyset$ **then goto** *loop*;
  end if
end of the procedure:
  return p^e;
end procedure

and its duration d_{jk}. Such a rule corresponds to insert p^{ne} in the OR session with the maximum unused OR time in such a way to minimize the overtime demand when d_{jk} is greater than (3.7). The aim is to balance the workload among the OR sessions. If one non-elective patient is scheduled in the OR session (h,k) then the NEWF returns p^{ne}, otherwise it returns p^e.

The online algorithm for the RTM with elective and non-elective patients starts every time a surgery ends in an OR session (j,k) and the value of (3.7) is greater than d_{jk} and/or at least one non-elective patient is waiting to be operated on. Accordingly, when $Q^\tau = \emptyset$ only the former condition has to be checked and it is equivalent to the condition (3.4) defined for the RTM with only elective patients because $\mu_{jk} = 0$ and $L'_j = L^\tau_{jk}$. The algorithm for the RTM performs three steps, that is the resequencing, the NEWF and the overtime allocation.

Finally, we would like to remark that the amount of effective used overtime could slightly exceed the maximum overtime available Ω. It depends on whether the overtime is assigned basing the decision on the EOT since ROT is not available.

Under special circumstances, extra overtime can be required for the surgery completion but all the available overtime has been previously assigned. In this case, we assume to allow the surgery completion setting the parameter Ω equal to the effectively used overtime.

3.4 *Ex-Post* Approach: The Offline Solution

The online solution is characterized by the lack of knowledge about what might happen in the remaining of the planning horizon. This is due to the difference between estimated and real duration of a surgery, and to the unforeseeable arrivals of non-elective patients.

On the contrary, at the end of the planning horizon we have a complete information about what is happened. Thus is possible to evaluate what would be the optimal decisions to be taken assuming to know in advance all the information that are acquired over time by the online solution. In our case, such information includes the ROTs of the elective patients and the surgery demand of the non-elective patients, that is their amount, the ROTs and the day in which they must be operated on.

We denote this set of decisions as *offline solution*. Such a solution provides a significant contribution to evaluate the effectiveness of the online approach. In this section, first we provide a linear programming model to compute the optimal offline solution in the case of only elective patients. Then, we extend this model to take into account also the non-elective patients.

Figure 3.2 reports an example in which the difference between EOTs and ROTs caused the request of an amount of overtime. In the OR session $(1, k)$ the overtime has been allocated in order to operate on the last patient, while in the OR session $(2, k)$ the surgery of the patient with index 3 has been postponed.

To determine an offline solution in the case of dealing with only the elective patients, the only relevant decision is that of postponing the surgery interventions. Since the ROTs are known, we remark that any sequencing of the surgery planned into an OR session determines the same outcome. Thus the sequencing is not relevant for the offline solution. Let us introduce the following decision variables

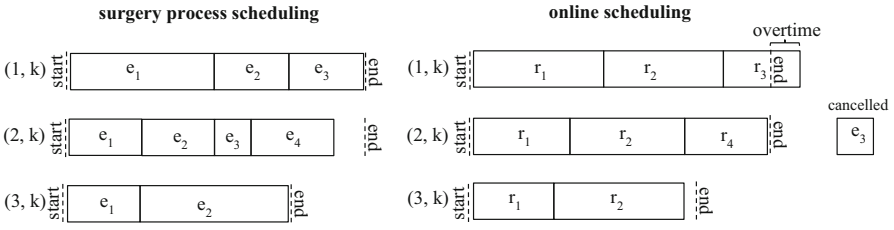

Fig. 3.2 Surgery process scheduling vs. online scheduling: elective patients

$$x_i = \begin{cases} 1 & \text{if the surgery intervention of the patient } i \in L \text{ is postponed} \\ 0 & \text{otherwise} \end{cases}$$

and the non-negative integer $v_{jk} \in \mathbb{Z}_+$ measuring the overtime assigned to the OR session (j,k).

To model the offline solution we introduce the following constraints:

$$\sum_{i \in L_{jk}} (1 - x_i) r_i \leq d_{jk} + v_{jk} \qquad \forall (j,k) \in S \tag{3.8}$$

$$\sum_{(j,k) \in S} v_{jk} \leq \Omega \tag{3.9}$$

$$x_i = 0 \qquad \forall i \in L_{\text{first}} \tag{3.10}$$

Constraints (3.8) ensure that the overall duration of the surgery performed during the OR session (j,k) can not exceed the duration of the OR session plus the additional overtime assigned. Constraint (3.9) limits the use of the overtime to the maximum overtime available. Finally, we remark that the first patient scheduled in each OR session (j,k) is not the subject of an online decision, that is he/she will be always operated on. Therefore, we are required to model this fact in our offline solution introducing the constraints (3.10) where $L_{\text{first}} \subset L$ is the set of all the patients sequenced as the first of an OR session.

We recall that our online solution would maximize the utilization of the OR sessions and to minimize the number of postponed patients whose $\widetilde{w} > 1$, that is those patients for which the MTBT will be exceeded. Thus our objective function should take into account these requirements.

We define the overall utilization of the OR sessions as the ratio between the total duration of the operated patients and the sum of the duration of all the OR sessions, limited to 1 to avoid greater values, that is when using the overtime

$$u = \min \left\{ \frac{\sum_{i \in L} (1 - x_i) r_i}{\sum_{(j,k) \in S} d_{jk}}, 1 \right\}.$$

To promote a solution with higher utilization, we introduce an auxiliary continuous variable $u \in [0,1]$ and the constraint

$$u \sum_{(j,k) \in S} d_{jk} \leq \sum_{i \in L} r_i (1 - x_i). \tag{3.11}$$

Our aim is to maximize the objective function defined as follows

$$z \equiv (1 - \alpha) u + \alpha \frac{\sum_{i \in L} (1 - x_i) - \sum_{i \in L_{\widetilde{w}>1}} x_i p_i}{|L|}, \tag{3.12}$$

which is the linear convex combination of two terms in $\alpha \in [0, 1]$. The former is the utilization defined by the constraint (3.11). The latter is the number of the patients operated on minus a sum of the penalties associated to those patient whose surgery is postponed and their $\widetilde{w} > 1$. Since the utilization ranges in $[0, 1]$, the latter term is normalized on the overall number of scheduled patient $|L|$. The penalties are defined as

$$p_i = \widetilde{w}_i^2 \,. \tag{3.13}$$

in order to limit the impact of the symmetries (see, e.g., Ghoniem and Sherali 2011).

Finally, the offline solution in the case of only elective patients can be computed by finding the optimal solution of the following mixed-integer linear program

$$\begin{aligned} M^e: \quad \max\, & z \quad \text{s.t.}(3.8) - (3.11) \\ & x_i \in \{0, 1\} \quad \forall i \in L \\ & v_{jk} \in \mathbb{Z}_+ \quad \forall (j, k) \in S \\ & u \in [0, 1] \end{aligned}$$

Figure 3.3 reports an example of solution of the problem in which two non-elective patients have been scheduled, determining the request of an amount of overtime for some of the OR sessions.

Model M^e can be modified to address also the management of the non-elective patients. Let Q_k be the set of the non-elective patient arrived the day k. We introduce the following decision variable

$$y_{ijk} = \begin{cases} 1 & \text{if the patient } i \in Q_k \text{ is inserted in the session } (j, k) \\ 0 & \text{otherwise} \end{cases} .$$

The constraints

$$\sum_{(j,k') \in S : k' = k} y_{ijk'} = 1 \quad \forall i \in Q_k,\ \forall k \in K \tag{3.14}$$

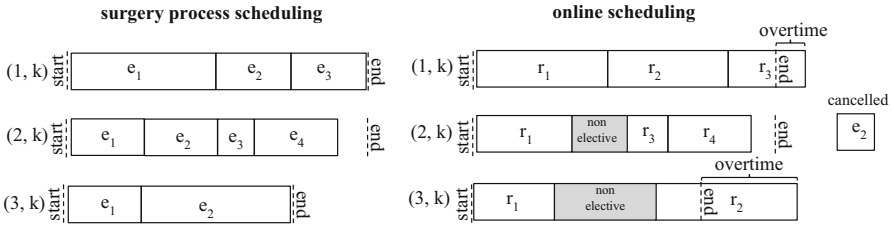

Fig. 3.3 Surgery process scheduling vs. online scheduling: elective and non-elective patients

ensure that each non-elective patient in Q_k is operated on during only one OR session (j,k). We remark that we have to modify the constraints (3.8) and (3.11) to take into account the insertion of the non-elective patients. By consequence, the new constraints are

$$\sum_{i \in L_{jk}} (1 - x_i) r_i + \sum_{i \in Q_k} y_{ijk} r_i \leq d_{jk} + v_{jk} \qquad \forall (j,k) \in S \tag{3.15}$$

$$u \sum_{(j,k) \in S} d_{jk} \leq \sum_{i \in L} (1 - x_i) r_i + \sum_{i \in Q} r_i \tag{3.16}$$

where $Q = \bigcup_{k \in K} Q_k$.

Finally, the offline solution in the case of elective and non-elective patients can be computed by finding the optimal solution of the following mixed-integer linear program

$$\begin{aligned} M^{ne}: \quad \max\, & z \quad \text{s.t. (3.9)-(3.10), (3.14)-(3.16)} \\ & x_i \in \{0,1\} \quad \forall i \in L \\ & v_{jk} \in \mathbb{Z}_+ \quad \forall (j,k) \in S \\ & y_{ijk} \in \{0,1\} \quad \forall i \in Q,\ \forall (j,k) \in S \\ & u \in [0,1] \end{aligned}$$

3.5 Quantitative Analysis

This section reports the quantitative analysis performed under several scenarios to evaluate the effectiveness of the proposed online methods providing two different but complementary analysis. The first one is to embed our online approaches on a simulated surgical clinical pathway in such a way to evaluate their impact on the RTM week by week, that is how the previous decisions (e.g., determining less or more cancellations) can impact on the current decisions. The second one exploits the computation of the corresponding offline solutions in such a way to assess the competitiveness of the proposed online solutions. The section is organized as follows. We describe the simulated surgical clinical pathway in Sect. 3.5.1 while we describe the performance indices and the different evaluation scenarios in Sect. 3.5.2. We report and discuss the results of our computational tests in Sect. 3.5.3 while some computational remarks are discussed in Sect. 3.5.4.

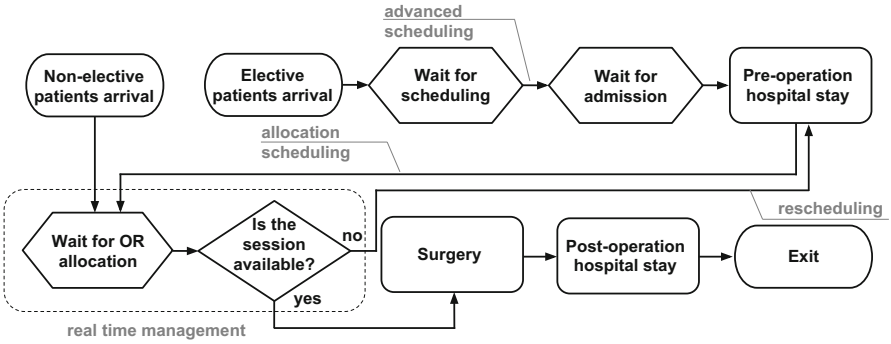

Fig. 3.4 The surgical CP and the optimization modules

3.5.1 The Simulated Surgical Clinical Pathway

In a surgical context, the main activities composing the clinical pathway are summarized in Fig. 3.4, which depicts also the moment on which the optimization problems arises. Our simulated surgical clinical pathway is a straightforward implementation of the surgical pathway described in Fig. 3.4. The simulation framework is based on a Discrete Event Simulation (DES) since it is the most suitable methodology to analyze a discrete and stochastic workflow. Further, DES is the only approach capable to represent the single entity within a CP, which is a necessary condition to apply our online algorithms.

The optimization modules embedded in the hybrid model are the RTM algorithms presented in Sect. 3.3 and the following:

Advanced scheduling: a metaheuristic based on a greedy construction of an initial solution and then a local search to improve that solution as reported in Duma and Aringhieri (2015);

Allocation scheduling: patients are sequenced in decreasing order of $\tilde{w}_i$.

Furthermore, the canceled surgeries are rescheduled in one of the OR sessions of the first day of the next week. We recall that the advanced scheduling is aimed at maximizing the OR utilization. This fact directly influences the number of possible cancellations during the scheduling posing a challenge for the RTM. Furthermore, we recall that it makes really difficult to insert a patient whose surgery has been postponed by the RTM, as reported in Duma and Aringhieri (2015). This justify our choice to schedule on the next week all the postponed patients.

The resulting hybrid model is implemented using AnyLogic 7.1 (Borshchev 2013): its Enterprise Library is exploited for the implementation of the DES simulation framework whilst the optimization modules are implemented from scratch in Java, which is the native programming language of AnyLogic.

Table 3.2 Parameters characterizing the four scenarios

Varying parameters					
Scenario	Non-elective	Ω	Scenario	Non-elective	Ω
(E)	–	10 h	(NE2)	30 per week	50 h
(NE1)	15 per week	15 h	(NE2b)	30 per week	40 h
Common parameters to all scenarios					
Parameter	Unit	Value	Parameter	Unit	Value
Elect. arrival rate	Patients/Day	25	Initial $\|I\|$	Patients	500
Avg. EOT	Minutes	140	S. dev. EOTs	Minutes	75
S. dev. $r_i - e_i$	Minutes	30	Max ROT	Minutes	480
n	Sessions/Week	45	d_{jk}	Minutes	480

Table 3.3 Urgency classes and MTBTs of the elective patients

URG class	Frequency	MTBT (days)	URG class	Frequency	MTBT (days)
A	3%	8	B	5%	15
C	7%	30	D	10%	60
E	15%	90	F	25%	120
G	35%	180			

3.5.2 Scenarios and Indices

The model briefly discussed in Sect. 3.5.1 is inspired to the case study reported in Ozcan et al. (2011) for the thyroid surgical treatment of elective patients. Since our model is capable to represent such a setting, in our previous work (Duma and Aringhieri 2015), we validated our model through a comparison with the one discussed in Ozcan et al. (2011). To avoid to get trapped on a single case study, which could be a limitation from our point of view, we introduce four scenarios in such a way to provide more accurate insights from our quantitative analysis. To this end, we will consider four scenarios (E, NE1, NE2, NE2b) obtained by varying the non-elective arrival ratio and available overtime while the other parameters characterizing them are fixed. All the parameters are reported in Table 3.2.

The flow of elective patients is described in terms of urgency class, frequency and MTBT in Table 3.3. Finally, all the simulation model parameters are the same of those reported in Duma and Aringhieri (2015).

Table 3.4 reports the two types of indices adopted to evaluate the impact of the optimization modules. We define a set of patient-centered indices in such a way to evaluate the performance from the patient point of view. We also define a set of facility-centered indices in such a way to evaluate them against to the patient-centered ones.

The reason of considering both patient- and facility-centered indices relies in the more general idea of using a CP to enhance the quality of care by improving patient outcomes, promoting patient safety, increasing patient satisfaction, and optimizing the use of resources.

Table 3.4 Patient-centered and facility-centered indices

Index	Definition
Patient-centered	
p	Number of surgeries performed
c	Number of cancellations
f_{MTBT}	Percentage of patients operated within the MTBT
ℓ_{avg}	Average length (number of patients) of the waiting list
t_{avg}	Average waiting time (days) spent in the waiting list
w_{avg}	Average value of patient's $w_i = t_i/t_i^{\max}$ at the time of their surgery
Facility-centered	
u_{OR}	OR utilization
u_{over}	Overtime utilization

3.5.3 Results

In this section we report the results of our quantitative analysis obtained by running our methods on the four different scenarios.

In order to provide a term of comparison, we introduce a baseline configuration in which the algorithms for the advanced scheduling and the allocation scheduling are those reported in Sect. 3.5.1 while the solution for the RTM is simpler than those proposed in Sect. 3.3. In the baseline configuration, the resequencing is not performed, the overtime is a priori uniformly distributed among the OR sessions and the non-elective patients are assigned to the first free OR session. The baseline configuration does not claim to fit perfectly the clinical practice (since we are not dealing with a specific case study) but it would represent a more general operative context in which some optimization approaches are performed on the planning side but without taking into account the inherent uncertainty arising in the management of a surgical pathway. The introduction of the baseline configuration allows us to evaluate the actual impact of the RTM on the management of the surgical pathway.

Two further configurations are introduced to properly evaluate the online approach in the case of non-elective patients, that is one configuration with the NEWF algorithm (conf. 2) and one without (conf. 1). When NEWF is not considered, the non-elective patients are scheduled as soon as possible.

Table 3.5 reports the value of the performance indices, which are obtained by taking the corresponding average value running the simulation model (depicted in Sect. 3.5.1) 30 times on a given configuration and, each time, starting from a different initial condition. Each run replicates 2 years of operating room management. Data are collected only on the second year. Remarks on running time are reported in Sect. 3.5.4.

With respect to the baseline configuration, the reported results showed that the adoption of the online approach for the RTM – both in the case of only elective or non-elective patients – can largely improves the patient-centered performance indices while maintaining the facility-centered ones.

Table 3.5 Performance indices for each scenario and RTM configuration

Scenario Id	RTM config.	Performance indices							
		p	c	f_{MTBT} (%)	ℓ_{avg}	t_{avg}	w_{avg}	u_{OR} (%)	u_{over} (%)
(E)	Baseline	7532	506	44.0	2879	111	1.05	83.4	13.3
	1	7991	399	91.9	2168	86	0.81	88.5	56.9
(NE1)	Baseline	7050	999	17.3	3400	130	1.25	86.5	17.5
	1	7657	693	65.6	2465	98	0.92	93.1	100.0
	2	7796	595	76.2	2317	93	0.87	94.6	100.0
(NE2)	Baseline	6866	1120	2.5	3855	147	1.42	93.9	21.2
	1	8131	366	96.6	1952	81	0.75	100.0	95.8
	2	8147	356	96.8	1921	80	0.74	100.0	96.2
(NE2b)	Baseline	6717	1244	0.2	4064	154	1.50	92.3	20.2
	1	7780	572	66.7	2440	99	0.92	100.0	98.0
	2	7878	540	77.4	2347	96	0.89	100.0	98.4

The most relevant improvement is that regarding the percentage of patients operated within the MTBT (f_{MTBT}), which measures the capability of the hospital to respect their deadlines ensuring to deliver a surgery in a proper way. The results prove the positive impact of the NEWF algorithm. For instance, an improvement of more than the 10% of the f_{MTBT} can be observed on the scenarios (NE1) and (NE2b) while this improvement is limited in the scenario (NE2).

Figure 3.5 reports the trend of the f_{MTBT} value during the simulation in order to compare the behavior in the scenario (NE2) (Fig. 3.5a) and (NE2b) (Fig. 3.5b). As reported in Table 3.2. the difference between the two scenarios is the amount of available overtime, that is 50 h for (NE2) and 40 for (NE2b). While in Fig. 3.5a the f_{MTBT} is quite stable along the time, we observe that in Fig. 3.5b a drop is reported after about 240 days. This highlights the fundamental role of the overtime as a flexible resources, when the required amount is correctly evaluated. We also remark that the NEWF is able to limit the negative impact of an overtime underestimation (Fig. 3.5b).

Table 3.6 reports the competitive analysis, that is the comparison between online and offline solutions. The results are obtained as follows. Among the 30 runs reported before, we selected the run whose performance indices are closest to the average values of the performance indices. From this run, we extracted the information required to generate the 52 instances corresponding to the second year of simulated time. Finally, we computed the optimal solution for each of the 52 instances by solving the corresponding mixed-integer linear problem.

Therefore, Table 3.6 reports the average values over 52 instances of the following quantities: π and π' are respectively the number of elective patients scheduled and the number of elective patients scheduled whose $\tilde{w}_i > 1$; γ and γ' are respectively the number of cancellations and the number of cancellations whose corresponding patients have $\tilde{w}_i > 1$; z_{avg} is the value of the objective function (3.12). Finally, the columns regarding the competitive ratio report both the experimental average and worst ratio.

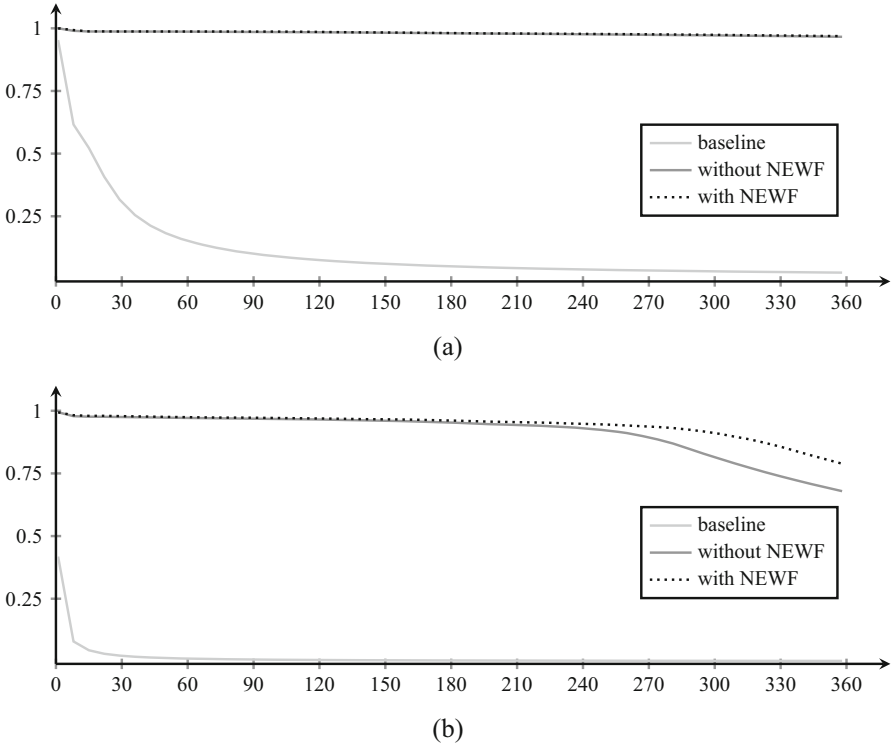

Fig. 3.5 Trend of f_{MTBT} (data referred to the second year, days on x-axis, percentage of patients on y-axis). Scenario (NE2) in (**a**). Scenario (NE2b) in (**b**)

The competitive analysis confirms the quality of the online solutions. In particular, the analysis of the z_{avg} values and the average and the worst competitive ratio values validates the remark about the positive impact of the NEWF algorithm.

The analysis of the average competitive ratio proves the challenging of the problem of dealing with the management of a flow of elective and non-elective patients. Actually, the competitive ratio of the baseline solution largely increases as soon as the arrival rate of the non-elective increases or the available overtime is tight. On the contrary, the competitive ratio of the configurations 1 and 2 is quite stable. Furthermore, the gap between the two competitive ratios (baseline vs. configuration 1 or 2) is quite acceptable for the scenario (E) while increases for the other non-elective scenarios demonstrating the need of an online solution to cope in a effective way the management of non-elective patients.

Table 3.6 Comparison between online and offline solutions

Scenario Id	RTM config.	Input	Online sol		Offline sol		Comp. ratio		Time (s)
		π (π')	γ (γ')	z_{avg}	γ (γ')	z_{avg}	Avg.	Worst	
(E)	Baseline	157 (78)	12 (5)	0.8607	3 (1)	0.9347	1.09	1.15	0.12
	1	161 (8)	7 (0)	0.9243	4 (0)	0.9467	1.02	1.09	0.11
(NE1)	Baseline	156 (109)	20 (13)	0.8022	2 (1)	0.9830	1.23	1.45	0.96
	1	161 (42)	13 (2)	0.9188	2 (0)	0.9883	1.08	1.26	35.31
	2	161 (24)	12 (0)	0.9353	2 (0)	0.9894	1.06	1.13	3.98
(NE2)	Baseline	154 (138)	22 (20)	0.7708	0 (0)	0.9978	1.30	1.63	0.20
	1	163 (6)	8 (0)	0.9768	1 (0)	0.9971	1.02	1.06	68.75
	2	163 (5)	7 (0)	0.9786	1 (0)	0.9972	1.02	1.06	0.53
(NE2b)	Baseline	152 (147)	23 (22)	0.7317	1 (1)	0.9929	1.37	1.78	0.49
	1	161 (37)	12 (1)	0.9616	2 (0)	0.9932	1.03	1.12	117.31
	2	161 (31)	10 (0)	0.9681	2 (0)	0.9939	1.03	1.07	110.16

3.5.4 Computational Remarks

The results reported in Sect. 3.5.3 are obtained running our computational tests on a 64 bit Intel Core i5 CPU with 3.33 GHz and 3.7 GB of main memory.

On average, one single run of the simulation model requires from 7.0 to 20.5 s when running with scenario (E) and baseline configuration or with scenario (NE2b) and configuration 2, respectively. This means that no more than 615 s are needed to simulate 2 years of operating room management. Finally, we remark that the algorithm for the advanced scheduling is the most time consuming component while the running time required by the other components is negligible.

The mixed-integer linear programs are solved using IBM ILOG CPLEX Optimization Studio 12.3. The CPLEX running time are reported in the last column of Table 3.6. Note that usually few seconds are enough to solve an instance of the offline problem. The high average values are determined by few instances requiring a lot of time to close the optimality gap. For example, the number of instances requiring more than 5 s in the scenario (NE2b) are 10 and 4 for configuration 1 and 2, respectively. This is probably due to the large number of symmetries determined by the decision variables y_{ijk} for a given day k.

3.6 Sharing Resources Among Surgical Pathways

The specific assignment of OR sessions to be shared among different surgical CPs, is defined by the Master Surgical Schedule (MSS). The MSS must be updated whenever the total amount of OR time changes or when the requirements of some surgical CPs change. This can occur not only as a response to long term changes in the overall OR capacity or staffing fluctuations, but also in response to seasonal

fluctuations in demand. As already discussed in the previous sections, RTM deals with the overtime management in the case of a single surgical CP. However, when the overtime is a shared resources, the online decision of using the overtime, or to cancel a surgery, should take into account a fairness criterion.

The objective of the management of the shared resources is therefore to have a fair assignment of the overtime and the OR sessions to surgical CPs.

In the following, we will consider two or more different surgical CPs corresponding to different specialties. Each CP is essentially the same surgical pathway described in Fig. 3.4 but without the source of non-elective patients. We will consider only elective patients since we would like to have a more clear idea of the impact of the proposed sharing policies: actually, in the current context, a flow of non-elective patients would correspond to a higher workload, that we are able to manage as shown in Sect. 3.5. For the sake of simplicity, we refer hereafter to those surgical CPs as specialties.

3.6.1 Policies for Sharing ORs

We would define how to assign the available OR sessions among different specialties in such a way to ensure enough and balanced OR sessions to each specialty. In our operative context, the MSS is updated every time period (usually 1 month or 1 week). To this end, we define three alternative policies.

The first policy is *Based on Lengths* (BL), that is, every 4 weeks, the OR sessions are reassigned so that they are proportional to the number of patients in the waiting list of each specialty. On the contrary, the second policy is *Based on the EOTs* (BE) in which every 4 weeks, the OR sessions are reassigned so that they are proportional to the sum of the EOTs of the patients in the waiting list of each specialty.

The last policy consists in a *Flexible Scheduling* (FS) in which MSS and Advanced Scheduling are solved at the same time every week. The algorithm implementing the FS policy is an adaptation of that proposed in Aringhieri et al. (2015). It consists of a greedy construction of the initial solution and an improvement phase performed by a local search engine: (1) at the beginning of the greedy construction, the patients are ordered by decreasing value of $\widetilde{w}_i$, and each OR session is not assigned to any specialty, except for the OR sessions used to reschedule the patients postponed during the last week; (2) during the greedy construction, patients can be inserted only into OR sessions assigned to their specialty, or into OR sessions not already assigned (that is empty OR sessions); in the latter case, such an OR session is assigned to the specialty of the patient; (3) during the local search, only swaps between patients belonging to the same specialty are allowed.

3.6.2 Policies for Sharing Overtime

When sharing overtime, we are interested in guaranteeing a fair access to the available overtime from the different specialties. We propose two alternative policies.

The first policy is called *Dedicated Overtime Allocation* (DOA). in which a dedicated amount Ω^s of weekly overtime is allocated to the specialty s, so that it is proportional to the number of OR sessions assigned by the MSS. By consequence, the RTM will take into account Ω^s as the overtime available when applying the criterion (3.6).

The second policy is called *Flexible Overtime Allocation* (FOA), in which all the specialties share the total available overtime Ω; in order to foster a balanced use of the overtime, we adapt the criterion (3.6) as follow:

$$\beta_{jk}^{\tau}\left(\frac{e_i+\rho_{jk}^{\tau}}{d_{jk}}\right)\left(1+\frac{v^s}{\Omega}-\frac{n_k^s}{n}\right)\leq 1 \tag{3.17}$$

where v^s is the amount of weekly overtime used by the specialty s until that time and n_k^s is the number of OR sessions of that specialty from the day after k.

The policy FOA introduces a new factor which measure the overtime still available with respect to the number of OR sessions to be still performed by the specialty. This value is closed to 1 when the overtime has been used proportionally with respect to the assigned and completed sessions. On the contrary, it is between 0 and 1 or it is greater than 1 when it is underused or overused, respectively.

3.6.3 Quantitative Analysis

Following the same analysis framework proposed in Sect. 3.5, we report a quantitative analysis to evaluate the impact of the policies for the resource sharing.

In the current analysis, we consider as "baseline" the configuration 1 introduced in Sect. 3.5.3, that is the best one when dealing with only elective patients. The only difference relies on the allocation scheduling in which the patients having $\tilde{w}_i \leq 1$ are sequenced with the Longest Processing Time (LPT) rule. This is due to the slightly more sophisticated policy used in Duma and Aringhieri (2015) and Aringhieri and Duma (2016), that would make it difficult to implement a linear programming model for the competitive analysis.

In the baseline configuration, the overtime is allocated using the DOA rule while the number of OR sessions are proportional to the arrival rate and it does not change over time.

Table 3.7 describes the two different scenarios in which we evaluate our sharing policies. The two scenarios differ from (1) the number of specialties, (2) the amount of available resources (number of operating rooms and weekly overtime), and (3) the patient features, namely the arrival rates, the EOT distributions and the urgency distributions.

Table 3.7 Parameters of the two scenario

Parameters	Unit of measure	Scenario S_1	Scenario S_2
Arrival rate			
Pathways 1, 2, 3	Patients/Day	12.5, 12.5, –	24.0, 12.0, 4.0
Initial waiting list	Patients	1000	1500
MTBT URG $A, \cdots, G$	Days	8, 15, 30, 60, 90, 120, 180	8, 15, 30, 60, 90, 120, 180
Frequency URG $A, \cdots, G$			
Pathway 1		5%, 15%, 40%, 15%, 10%, 10%, 5%	5%, 15%, 40%, 15%, 10%, 10%, 5%
Pathway 2		14%, 14%, 14%, 14%, 15%, 15%, 15%	14%, 14%, 14%, 14%, 15%, 15%, 15%
Pathway 3			8%, 9%, 11%, 12%, 15%, 15%, 30%
EOT average			
Pathways 1,2,3	Minutes	120, 180, –	150, 120, 180
EOT deviation			
Pathways 1,2,3	Minutes	75, 75, –	75, 75, 75
n	ORs	50 (10 a day)	75 (15 a day)
$\mathbf{d_{jk}}$	Minutes	480	480
Ω	Minutes	600	900

Table 3.8 Scenario S_1: performance indices (B is the baseline configuration)

	Policies		Performance indices							
Id	MSS	FOA	p	c	f_{MTBT} (%)	ℓ_{avg}	t_{avg}	w_{avg}	u_{OR} (%)	u_{over} (%)
B			8380	352	56	1634	56	0.99	88	44
1	BL		9058	387	91	984	43	0.75	97	77
2	BE		9040	380	87	973	44	0.79	97	74
3	FS		9050	380	93	918	42	0.74	97	79
4	BL	√	8903	539	81	1134	50	0.87	95	77
5	BE	√	8895	493	61	1215	53	0.94	95	59
6	FS	√	8850	628	67	1229	53	0.94	95	69

Table 3.9 Scenario S_2: performance indices (B is the baseline configuration)

	Policies		Performance indices							
Id	MSS	FOA	p	c	f_{MTBT} (%)	ℓ_{avg}	t_{avg}	w_{avg}	u_{OR} (%)	u_{over} (%)
B			13547	609	33	2410	56	1.09	92	26
1	BL		14189	626	58	1853	51	0.91	97	52
2	BE		14198	605	60	1877	51	0.92	97	54
3	FS		14238	577	83	1864	51	0.88	97	74
4	BL	√	14031	725	40	2145	57	1.03	96	46
5	BE	√	14009	718	38	2179	58	1.04	96	45
6	FS	√	13858	870	39	2289	60	1.05	95	46

Tables 3.8 and 3.9 report the analysis of the proposed policies for the resource sharing. We denote with "B" the baseline configuration while those obtained by combining the different policies are denoted with an integer from 1 to 6. The combination of the different policies is described in the second and the third columns: the column "MSS" denotes the policy used for OR sharing while the column "FOA" indicates when the FOA policy has been adopted in alternative to the DOA one. Finally, for each configuration, the performance indices are reported.

Considering the scenario S_1 in Table 3.8, we can remark that all the configurations 1–6 indicates a general improvement of the performance indices with respect to the baseline configuration, except for the number of cancellations. This justify the need of ad hoc solutions to deal with the resource sharing. Furthermore, the DOA rule seems to be more effective than the FOA, especially when considering the patient-centered indices. These considerations are confirmed also by the analysis reported in Table 3.9 for the Scenario S_2, in which the effectiveness of the DOA with respect to the FOA is more evident.

Considering both scenarios S_1 and S_2, the configuration 3 – that is, that adopting a flexible scheduling policy for the MSS and a dedicated allocation for the overtime – leads to a general and robust improvement of all the indices resulting as the best configuration. Table 3.10 reports in more detailed way the results for that configuration reporting also the performance indices for the two pathways analyzed in the scenario S_1.

Table 3.10 Configuration 3, scenario S_1: detailed analysis

Id	Pathways	Performance indices							
		p	c	f_{MTBT} (%)	ℓ_{avg}	t_{avg}	w_{avg}	u_{OR} (%)	u_{over} (%)
B	Both	8380	352	56	1634	56	0.99	88	44
	1	4587	251	99	40	8	0.25	80	77
	2	3793	101	5	1 595	114	1.89	96	11
3	Both	9050	380	93	918	42	0.74	97	79
	1	4544	289	94	392	36	0.75	98	92
	2	4507	91	91	526	47	0.73	97	70

From the results for each pathway reported in Table 3.10, it is evident the effectiveness of the flexible scheduling policy: actually, the results for configuration 3 demonstrate a good balance of the performance indices relative to the two pathways, especially that regarding the percentage of surgeries performed within their MTBT threshold.

3.7 Conclusions

The RTM of operating rooms is the decision problem arising during the fulfillment of the surgery process scheduling, that is the problem of supervising the execution of such a schedule and, in case of delays, to take the more rational decision regarding the surgery cancellation or the overtime assignment. The RTM is characterized by the uncertainty of its main parameters. In this chapter, we considered two challenging extensions of the original problem with only elective patients. The resulting quantitative analysis showed the crucial role of the RTM of the operating rooms.

The first extension is that of considering a joint flow of elective and non-elective patients. We evaluated the effectiveness of the RTM on a simulated surgical clinical pathway under several scenarios and also reporting a competitive analysis with respect to an offline solution obtained by solving a mixed-integer programming model. The quantitative analysis showed the capability of the online solutions to address the inherent uncertainty of the RTM determining a general improvement of the patient-centered indices without deteriorating the facility-centered ones. Further, the analysis of the competitive ratios confirmed the challenging of the problem of dealing with a flow of non-elective patients sharing the ORs with a flow of elective patients.

The second extension dealt with the management of the shared resources among different surgical pathways. The shared resources considered in our analysis are the ORs and the overtime. Our analysis demonstrated that different pathways can benefit from sharing the resources when adequate policies are adopted.

From a methodological point of view, our analysis suggested that online optimization can be a suitable methodology to deal with the inherent stochastic aspects arising in the majority of the health care problems. Although online optimization does not exploit sophisticated mathematical approaches, the competitive analysis reported in Table 3.6 suggested its capability to deal with the stochastic aspects of a problem whenever such aspects are embedded into a well-structured optimization problem, such as those arising in the health care management.

Future research avenues could consider a more systematic analysis of the optimization solutions provided in literature to deal with non-elective patients under the sharing OR policy, and to compare the results with those obtained by the dedicated OR policy.

Acknowledgements The authors acknowledge support from the Italian Ministry of Education, University and Research (MIUR), under the grant n. RBFR08IKSB, "Firb—Futuro in Ricerca 2008".

References

Addis B, Carello G, Grosso A, Tànfani E (2016) Operating room scheduling and rescheduling: a rolling horizon approach. Flex Serv Manuf J 28(1–2):206–232

Aringhieri R, Duma D (2016) The optimization of a surgical clinical pathway. In Obaidat MS et al (eds) Simulation and modeling methodologies, technologies and applications. Advances in intelligent systems and computing, vol 402. Springer, Cham, pp 313–331.

Aringhieri R, Tànfani E, Testi A (2013) Operations research for health care delivery. Comput Oper Res 40(9):2165–2166

Aringhieri R, Landa P, Soriano P, Tànfani E, Testi A (2015) A two level metaheuristic for the operating room scheduling and assignment problem. Comput Oper Res 54:21–34

Banditori C, Cappanera P, Visintin F (2013) A combined optimization-simulation approach to the master surgical scheduling problem. J Manag Math 24:155–187

Beaulieu I, Gendreau M, Soriano P (2012) Operating rooms scheduling under uncertainty. In: Tànfani E, Testi A (eds) Advanced decision making methods applied to health care. International series in operations research and management science, vol 173. Springer, Milan, pp 13–32. ISBN 978-88-470-2320-8

Borshchev A (2013) The big book of simulation modeling. Multimethod modeling with anyLogic. AnyLogic North America, Hampton, NJ. ISBN 978-0-9895731-7-7

Bruni M, Beraldi P, Conforti D (2015) A stochastic programming approach for operating theatre scheduling under uncertainty. IMA J Manag Math 26(1):99–119

Campbell H, Bradshaw N, Porteous M (1998) Integrated care pathways. Br Med J 316:133–144

Cardoen B, Demeulemeester E, Beliën J (2010) Operating room planning and scheduling: a literature review. Eur J Oper Res 201:921–932

Duma D, Aringhieri R (2015) An online optimization approach for the real time management of operating rooms. Oper Res Health Care 7:40–51

Erdem E, Qu X, Shi J, Upadhyaya S (2011) A mathematical modeling approach for rescheduling elective admissions upon arrival of emergency patients. In: 61st annual IIE conference and expo proceedings

Erdem E, Qu X, Shi J (2012) Rescheduling of elective patients upon the arrival of emergency patients. Decis Support Syst 54(1):551–563

Ghoniem A, Sherali H (2011) Defeating symmetry in combinatorial optimization via objective perturbations and hierarchical constraints. IIE Trans (Institute of Industrial Engineers) 43(8):575–588

Guerriero F, Guido R (2011) Operational research in the management of the operating theatre: a survey. Health Care Manag Sci 14:89–114

Hans EW, Vanberkel PT (2012) Operating theatre planning and scheduling. In: Hall R (ed) Handbook of healthcare system scheduling. International series in operations research and management science, vol 168. Springer, pp 105–130. ISBN 978-1-4614-1733-0

Heng M, Wright JG (2013) Dedicated operating room for emergency surgery improves access and efficiency. Can J Surg 56(3):167–174

Landa P, Aringhieri R, Soriano P, Tànfani E, Testi A (2016) A hybrid optimization algorithm for surgeries scheduling. Oper Res Health Care 8:103–114

Martello S, Toth P (1990) Knapsack problems: algorithms and computer implementations. Wiley-interscience series in discrete mathematics and optimization. Wiley, Chichester

Ozcan Y, Tànfani E, Testi A (2011) A simulation-based modeling framework to deal woth clinical pathways. In: Jain S, Creasey R, Himmelspach J, White K, Fu M (eds) Proceedings of the 2011 winter simulation conference, pp 1190–1201

Pham D-N, Klinkert A (2008) Surgical case scheduling as a generalized job shop scheduling problem. Eur J Oper Res 185(3):1011–1025

Stuart K, Kozan E (2012) Reactive scheduling model for the operating theatre. Flex Serv Manuf J 24(4):400–421

Testi A, Tànfani E (2009) Tactical and operational decisions for operating room planning: efficiency and welfare implication. Health Care Manag Sci 12:363–373

Testi A, Tànfani E, Torre G (2007) A three-phase approach for operating theatre schedules. Health Care Manag Sci 10:163–172

van Oostrum J, van Houdenhoven M, Hurink J, Hans E, Wullink G, Kazemier G (2008) A master surgical scheduling approach for cyclic scheduling in operating room departments. OR Spectrum 30:355–374

van Oostrum J, Bredenhoff E, Hans E (2010) Suitability and managerial implications of a master surgical scheduling approach. Ann Oper Res 178(1):91–104

Van Riet C, Demeulemeester E (2015) Trade-offs in operating room planning for electives and emergencies: a review. Oper Res Health Care 7:52–69

van Veen-Berkx E, van Dijk M, Cornelisse D, Kazemier G, Mokken F (2016) Scheduling anesthesia time reduces case cancellations and improves operating room workflow in a university hospital setting. J Am Coll Surg 223(2):343–51

Wullink G, Van Houdenhoven M, Hans EW, Van Oostrum JM, Van Der Lans M, Kazemier G (2007) Closing emergency operating rooms improves efficiency. J Med Syst 31(6):543–546

Chapter 4
Mixed Fuzzy Clustering for Deriving Predictive Models in Intensive Care Units

Cátia M. Salgado, Susana M. Vieira, and João M.C. Sousa

4.1 Introduction

Intensive care unit (ICU) data has grown exponentially in the last decades, making the ICU a particularly appealing setting for the implementation of data-based systems (Celi et al. 2013). Such systems acquire large quantities of data to discover hidden associations and understand patterns and trends in data, which can be used for diagnostic, prognostic and therapy (Celi et al. 2013; Vieira et al. 2013).

ICU databases contain records of patients' vital signs, laboratory results, prescribed and administrated medications, fluid balance, nursing notes, imaging reports, demographic information and clinical history. The premise of this work is that the complex and non linear relationship between different types of variables should be investigated and accounted for when deriving predictive models to support medical decision making. However, the mixed nature of data in electronic medical records still poses a daunting challenge for deriving data-based predictive models in the ICU. In particular, the available options to simultaneously handle static and time variant data are limited, and traditional methods do not provide the required means to extract useful knowledge that accounts for correlations between both.

In Izakian et al. (2013), the fuzzy c-means clustering technique is augmented for spatiotemporal clustering, a form of grouping objects based on their spatial (static) and temporal similarity. We extend this algorithm to any data set containing both time variant and time invariant features (mixed datasets) of any size, and use the knowledge extracted from the identified mixed structures to derive fuzzy models,

C.M. Salgado (✉) • S.M. Vieira • J.M.C. Sousa
IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Avenida Rovisco Pais, Lisboa, Portugal
e-mail: catia.salgado@tecnico.ulisboa.pt; susana.vieira@tecnico.ulisboa.pt; jmsousa@tecnico.ulisboa.pt

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_4

following the work of Ferreira et al. (2015). Fuzzy modeling can provide transparent predictive models and linguistic interpretations of the decision making process, showing great potential in dealing with vague information. Hence, it is especially well suited for health care applications since it can provide clinical insight into the classifier structure. We propose two different fuzzy modeling strategies based on the mixed fuzzy clustering algorithm and compare it with similar strategies based on fuzzy c-means. We show the results in four health care applications, for the prediction of adverse events of critically ill patients, including medication administration, mortality and readmissions.

The next section presents an overview of ICU adverse events. The mixed fuzzy clustering algorithm and the proposed modeling strategies are presented in Sects. 4.3 and 4.4, respectively. Data description and pre-processing is presented in Sect. 4.5, while experimental results for each application are shown in Sect. 4.6. Main conclusions and future work are presented in Sect. 4.7.

4.2 Adverse Events in the ICU

Several studies have focused on determining the predictive value of different types of numerical information in the prediction of adverse events, for improving the outcome of patients in the ICU. In particular, studies have focused on the prediction of mortality (Badawi and Breslow 2012; Clermont et al. 2001; Daly et al. 2001; Frize et al. 2001; Hug and Szolovits 2009; Ouanes et al. 2012; Reini et al. 2012; Ferreira et al. 2015), readmissions (Badawi and Breslow 2012; Campbell et al. 2008; Fialho et al. 2013, 2012; Frost et al. 2010; Gajic et al. 2008; Ouanes et al. 2012; Strand and Flaatten 2008; Walsh and Hripcsak 2014; Zheng et al. 2015; Ferreira et al. 2015) and length of stay (Frize et al. 2001; Reini et al. 2012).

Scoring systems in clinical use in the ICU include the Modified Early Warning Score (MEWS) (Reini et al. 2012), the Stability and Workload Index for Transfer (SWIFT) (Gajic et al. 2008), the Simplified Acute Physiology Score (SAPS II) (Le Gall et al. 1993), the Acute Physiology and Chronic Health Evaluation (APACHE II) (Knaus et al. 1985) and the Sequential Organ Failure Assessment (SOFA) (Vincent et al. 1996). MEWS attributes a score to each patient based on their respiratory, circulatory and neurological state, renal function and body temperature. SWIFT has been specifically developed to predict readmissions to the ICU; it represents a measure of five parameters: ICU admission source, length of ICU stay, respiratory impairment and neurological status, using the Glasgow Coma Scale (GCS) at discharge. APACHE II gives a score of the severity of illness; it is based on the patient's age and punctual physiological measurements of temperature, mean arterial pressure, pH, heart rate, respiratory rate, sodium, potassium, creatinine, hematocrit, white blood cells counts, GCS and partial pressure of arterial oxygen (PaO2), taken during the first 24 h after admission.

Currently, there has been an attempt to improve these conventional standard logistic regression techniques using machine learning algorithms such as artificial

neural networks, fuzzy logic and decision trees, which resulted in a number of predictive models in different ICUs (Fialho et al. 2013, 2012; Walsh and Hripcsak 2014; Zheng et al. 2015; Salgado et al. 2016). In spite of the growing popularity of these models among the research community, the role they play in supporting the physicians' decisions and in improving patients' outcomes remains uncertain (Allaudeen et al. 2011; Kansagara et al. 2011; Ouanes et al. 2012).

To the best of the authors knowledge, the simultaneous mining of time series and time invariant data has not been addressed in any of the aforementioned studies, which suggests that information about the patients may be being lost in the data mining process. Hence, we developed predictive models based on MFC using information about variables with more than one record over the patient' stay, including vital signs and laboratory results, and information that remains static throughout the stay, such as gender, weight and admission status. Two main cohorts of patients are considered: patients in septic shock and patients that were readmitted within 24–72 h of discharge.

4.2.1 Septic Shock: Vasopressors Administration and Mortality

Sepsis is a systemic whole-body inflammatory response to infection. Septic shock is an outcome of a sepsis reaction, associated with multiorgan failure and out of the normal range measurements of blood pressure, temperature, respiratory rate and white blood cells counts. It is one of the most common reasons of death in intensive care units, with a mortality rate of about 50%. A patient is considered to be in septic shock when the hypotensive state related to a sepsis condition persists, causing severe malfunction of vital organs despite adequate fluid resuscitation.

The initial priority in managing septic shock is to restore blood pressure and cardiac output by intravenous fluids administration. When fluid resuscitation is unable to restore an adequate arterial pressure, therapy with vasopressors must be initiated. Vasopressors are hypotension blood vessels drugs that are very effective in increasing blood pressure. The procedure of vasopressors administration is risky, since the catheter insertion involves a surgical procedure that can be associated with infections. These complications are increased when the procedure is done urgently such as in the case of unexpected systemic shock. Knowing beforehand which patients will need vasopressors would reduce the number of times the procedure is implemented and reduce the associated risk of infection, since clinicians would have more time to safely initiate the central line insertion protocol. This would in turn substantially improve the outcomes of these patients.

When the septic shock is caused mainly by abdominal indications it is called abdominal septic shock. Previous works have applied knowledge-based neural networks (Paetz 2003) and neuro-fuzzy techniques for predicting the outcome of these patients. In Fialho et al. (2010), the authors implement ant colony and bottom-up tree search techniques, combined with fuzzy modeling and neural networks, to determine the set of features more correlated with the mortality of these patients.

This study attempts to predict vasopressors administration and mortality of ICU patients in abdominal septic shock.

4.2.2 Early Readmissions

Patients readmitted to the ICU have increased mortality, morbidity and length of stay. ICU readmissions are regarded as an indicator of poor care and represent increased costs to the hospital (Allaudeen et al. 2011; Boudesteijn et al. 2007). According to Elliott et al. (2014), ICU readmission rates reported in literature vary between 1.3 and 13.7%. Although patients can have an unplanned readmission for any reason, from incomplete treatment or poor care to poor coordination of services at the time of discharge and afterwards, many of these readmissions are potentially preventable (Goldfield et al. 2008). Adequate risk stratification at the time of discharge could reduce readmission rates and hence improve patient outcomes.

Risk factors for ICU readmission have been systematically reported in prospective and retrospective cohort studies. The most commonly identified factors include: patient location before ICU admission, SAPS II and APACHE II scores at admission, age, co-morbidities, ICU length of stay, physiologic abnormalities at the time of ICU discharge or on the ward, ICU discharge at night or after hours, discharge to another critical care area or hospital, shock index (heart rate/systolic blood pressure), respiratory rate, Glasgow Coma Score, and higher Nursing Activity Score at the time of discharge (Rosenberg and Watts 2000; Rosenberg et al. 2001).

4.3 Mixed Fuzzy Clustering

Mixed fuzzy clustering (MFC) is a novel clustering method based on fuzzy c-means (Bezdek et al. 1984) which deals with both time variant and time invariant features (Ferreira et al. 2015). This method introduces a generalization of the spatiotemporal concept to any set of time variant and time invariant features and its extension to the analysis of multiple time series. Each entity x_i, with $i = 1, \ldots, N$, is characterized by features that are constant during the sampling time in analysis, $\mathbf{x}_i^s$, and by features that change over time (multiple time series), X_i^t:

$$x_i = (\mathbf{x}_i^s, X_i^t) \tag{4.1}$$

The time invariant component of the entities is represented by $\mathbf{x}_i^s$, where R is the number of time invariant features:

$$\mathbf{x}_i^s = (x_{i1}^s, \ldots, x_{iR}^s) \tag{4.2}$$

In order to extend the spatiotemporal clustering method proposed in Izakian et al. (2013), which only deals with one time series, to the case of multiple time series, a new dimension is introduced to handle P time variant features. The time variant component of the entities is represented by the matrix X_i^t:

$$X_i^t = \begin{pmatrix} x_{i11}^t & x_{i12}^t & \cdots & x_{i1P}^t \\ x_{i21}^t & x_{i22}^t & \cdots & x_{i2P}^t \\ \vdots & \vdots & \ddots & \vdots \\ x_{iQ1}^t & x_{iQ2}^t & \cdots & x_{iQP}^t \end{pmatrix}, \tag{4.3}$$

where Q is the length of time series.

The time invariant prototypes $\mathbf{v}_l^s$ for each cluster l are given by:

$$\mathbf{v}_l^s = \frac{\sum_{i=1}^N u_{li}^m \mathbf{x}_i^s}{\sum_{i=1}^N u_{li}^m}, \tag{4.4}$$

where $l = 1, \ldots, C$. The time variant prototypes $\mathbf{v}_{lk}^t$ for each cluster l and feature k are given by:

$$\mathbf{v}_{lk}^t = \frac{\sum_{i=1}^N u_{li}^m \mathbf{x}_{ik}^t}{\sum_{i=1}^N u_{li}^m} \tag{4.5}$$

The matrix of time variant prototypes for cluster l is represented by V_l^t.

In the above equations, the membership degree u_{li} of entity i to cluster l is given by:

$$u_{li} = \frac{1}{\sum_{o=1}^C \left(\frac{d_\lambda^2(\mathbf{v}_l^s, V_l^t, x_i)}{d_\lambda^2(\mathbf{v}_o^s, V_o^t, x_i)} \right)^{\frac{1}{m-1}}}, \tag{4.6}$$

where $m \in]1, \infty]$ is a weighting exponent that controls the degree of fuzziness.

The MFC algorithm clusters the data using an augmented form of the FCM. The main difference between the augmented and the classical FCM relies on the distance function. In the augmented FCM a new pondering element λ is included, factoring the importance to be given to the time variant component. The distance is also calculated separately for each time series. The distance function between a sample and the time invariant and time variant prototype of a cluster is computed:

$$d_\lambda^2(\mathbf{v}_l^s, V_l^t, x_i) = ||\mathbf{v}_l^s - \mathbf{x}_i^s||^2 + \lambda \sum_{k=1}^P ||\mathbf{v}_{lk}^t - \mathbf{x}_{ik}^t||^2 \tag{4.7}$$

The augmented FCM objective function is given by:

$$J = \sum_{l=1}^C \sum_{i=1}^N u_{li}^m d_\lambda^2(\mathbf{v}_l^s, V_l^t, x_i) \tag{4.8}$$

The MFC is described in Algorithm 1. Its inputs are the time invariant $\mathbf{x}^s$ and time variant data X^t, number of clusters C, initial partition matrix $U = [u_{li}]$, degree of fuzziness m and time variant component weight λ. It returns the final partition matrix $U = [u_{li}]$ and the time invariant V^s and time variant V^t prototypes. J^n represents the objective function in iteration n.

Algorithm 1 Mixed fuzzy clustering (MFC)

Input:
C: number of cluster prototypes
m: degree of fuzziness
λ: time variant component weight
$\mathbf{x}^s$: $N \times R$ matrix of time invariant data
X^t: $N \times Q \times P$ matrix of time variant data
$\mathbf{U}$: $C \times N$ random initial partition matrix
ϵ: stop criterion
Output:
$\mathbf{U}$: $C \times N$ partition matrix
$\mathbf{V}^s$: $C \times R$ matrix of time invariant cluster prototypes
V^t: $C \times Q \times P$ matrix of time variant cluster prototypes
while $\Delta J > \epsilon$ **do**
    Compute the cluster prototypes $\mathbf{v}_l^s$
    for k in $\{1, \ldots, P\}$ **do**
        Compute the cluster prototypes $\mathbf{v}_{lk}^t$
    end for
    Compute the distances $d_\lambda^2(\mathbf{v}_l^s, V_l^t, x_i)$
    Update the partition matrix $\mathbf{U}$
    Compute $\Delta J = J^n - J^{n-1}$
    $n = n + 1$
end while

4.4 Fuzzy Modeling Based on Mixed Fuzzy Clustering

4.4.1 Takagi-Sugeno Fuzzy Modeling

Fuzzy models are "grey box" and transparent models that allow the approximation of non linear systems with no previous knowledge of the system to be modeled. Fuzzy models have the advantage of not only providing transparency, but also linguistic interpretation in the form of rules.

In this work, Takagi-Sugeno (TS) fuzzy models (FMs) (Takagi and Sugeno 1985) are derived from data. These consist of fuzzy rules describing a local input-output relation. With TS FM, each discriminant function consists of rules of the type:

$$R_i : \text{If } x_1 \text{ is } A_{i1} \text{ and } \ldots \text{ and } x_M \text{ is } A_{iM}$$
$$\text{then } y(\mathbf{x}) = f_i(\mathbf{x}), \quad i = 1, 2, \ldots, C, \tag{4.9}$$

where A_{ij} are the antecedent fuzzy sets, f_i is the consequent function of rule R_i and y is the output. The degree of activation of the ith rule is given by $\beta_i = \prod_{j=1}^{M} \mu_{A_{ij}}(\mathbf{x})$, where $\mu_{A_{ij}}(\mathbf{x}) : \mathbb{R} \rightarrow [0, 1]$. The output is computed by aggregating the individual rules contributions:

$$y(\mathbf{x}) = \frac{\sum_{i=1}^{C} \beta_i f_i(\mathbf{x})}{\sum_{i=1}^{C} \beta_i} \tag{4.10}$$

The number of rules C and the antecedent fuzzy sets A_{ij} are determined by fuzzy clustering in the space of the input and output variables. The consequent functions $f_i(\mathbf{x})$ are linear functions determined by ordinary least squares.

For classification, a threshold γ is required to turn the continuous output $y \in [0, 1]$ into the binary output $y \in \{0, 1\}$. This way, an entity $\mathbf{x}$ is labeled as 1 if $y(\mathbf{x}) \geqslant \gamma$ and as 0 otherwise.

4.4.2 Proposed TS Fuzzy Models

Distinct Takagi Sugeno fuzzy model approaches based on clustering were considered for this study. The strategies differ in the type of input data and in the methodology used to determine the antecedent fuzzy sets. In particular, the antecedent fuzzy sets and the number of rules of the TS fuzzy model are determined based either on the partition matrix generated by the FCM algorithm (FCM FM), or in the partition matrix generated by MFC (MFC FM). MFC FM methodology was developed based on the belief that the identification of the fuzzy membership functions should be based on a non-conventional clustering algorithm, in the presence of a mix of time variant and time invariant features. Time variant features should not be directly mixed with time invariant features when calculating distances, and different time variant features should also be dealt with separately.

The input variables consist of (1) time variant and time invariant features or (2) transpose of the partition matrix generated by MFC ($\mathrm{U}^{\mathrm{MFC}}$). When time variant and time invariant data are used as input for the FCM FM, each time stamp of the time series is treated as one feature, i.e., the input of the model consists of a $N \times (R + Q \times P)$ matrix. When using the partition matrix, each feature corresponds to the degree of membership of the entities to the clusters such that the number of features equals the number of clusters determined in the clustering step. In particular, the input becomes u_{il}, where $i = 1, 2, \ldots, N$ and $l = 1, 2, \ldots, C$, which corresponds to the transpose of the partition matrix in (4.6). This approach can be seen as a type of feature transformation method.

Two approaches based on MFC are presented: MFC FM and FCM–$\mathrm{U}^{\mathrm{MFC}}$ FM, and compare it with the traditional FCM FM. The approaches listed below are described in Algorithm 2.

- FCM FM: Antecedent fuzzy sets determined by FCM in the space of the input and output variables.
- MFC FM: Antecedent fuzzy sets determined by MFC in the space of the input and output variables.
- FCM–U^{MFC} FM: Antecedent fuzzy sets determined by FCM in the space of the partition matrix generated by MFC and output variable.

4.5 Data Description

This paper uses two de-identified publicly available ICU databases, MIMIC II and MEDAN, which are described in the following.

The MIMIC II (Multi-parameter Intelligent Monitoring for Intensive Care) database is an ICU database from the Beth Israel Deaconess Medical Center in the United States (Saeed et al. 2011). Version 2.6 used in this study contains demographics, medications, laboratory results and other clinical data from 32,535 patients, collected over a 7-year period. Three datasets were built for classification using clinical, demographic and score information of adult patients (>15 years old at time of admission).

The MEDAN database (Hanisch et al. 2003) contains data from patients under abdominal septic shock registered from 1998 to 2002 by medical documentation staff at 71 ICUs in Germany. This dataset contains demographics and measurements of physiological variables from 410 patients, collected during their stay in the ICU.

4.5.1 Data Processing

Medical datasets are typically very heterogeneous (Paetz et al. 2004), due to its multiple and sometimes dissimilar sources. Each patient has different periods of time staying in medical facilities, during which distinct variables are documented. In addition, equipment and human faults, as well as seldom measurements are frequent. Particularly for retrospective evaluations, the quality of results relies heavily on the preprocessing of original data. Problems commonly associated with these datasets are the existence of missing data, uneven sampling times and outliers.

Missing data is a common occurrence in ICU databases either due to intentional reasons, i.e. data is irrelevant for the clinical problem under consideration and thus is not recorded, or unintentional reasons, when some kind of intervention or activity renders the data useless. In this work, patients and variables were initially selected in order to minimize missing data. When recoverable, missing data was filled using the zero order hold procedure, while unrecoverable data led to patient discarding.

Algorithm 2 Takagi-Sugeno fuzzy models

Data:
$\mathbf{x}^s$: $N \times R$ matrix of time invariant data
X^t: $N \times Q \times P$ matrix of time variant data
$\mathbf{Y}$: $N \times 1$ vector of class labels
$[\mathbf{x}^s \parallel X^t]$: matrix of concatenated input data
$[\mathbf{x}^s \parallel X^t \parallel \mathbf{Y}]$: matrix of concatenated input and output data
Parameters:
C: number of cluster prototypes
$\mathbf{U}$: $C \times N$ initial partition matrix
m: degree of fuzziness
λ: time variant component weight
$\mathrm{U}^{\mathrm{MFC}}$: $C \times N$ final MFC partition matrix
$(\mathrm{U}^{\mathrm{MFC}})^T$: $N \times C$ transpose matrix of $\mathrm{U}^{\mathrm{MFC}}$
$\mathrm{U}^{\mathrm{FCM}}$: $C \times N$ final FCM partition matrix
function *trainFCM*:($[\mathbf{x}^s \parallel X^t \parallel \mathbf{Y}]$, C, m)
    train a TS FM using product-space FCM clustering
    return model
end function
function *trainMFC*:($[\mathbf{x}^s \parallel X^t \parallel \mathbf{Y}]$, C, m, λ, U^{MFC})
    train a TS FM using product-space MFC clustering
    return model
end function
procedure FCM FM
    $[\mathrm{U}^{\mathrm{FCM}}, V]$ = FCM($[\mathbf{x}^s \parallel X^t \parallel \mathbf{Y}]$, C, $\mathbf{U}$, m)
    model=trainFCM($[\mathbf{x}^s \parallel X^t \parallel \mathbf{Y}]$, C, m)
    Y_m=test($[\mathbf{x}^s \parallel X^t]$, model)
end procedure
procedure MFC FM
    $[\mathrm{U}^{\mathrm{MFC}}, \mathbf{v}_l^s, V_l^t]$ = MFC($[\mathbf{x}^s \parallel \mathbf{Y}]$, X^t, C, $\mathbf{U}$, m, λ)
    model=trainMFC($[\mathbf{x}^s \parallel X^t \parallel \mathbf{Y}]$, C, m, $\mathrm{U}^{\mathrm{MFC}}$)
    Y_m=test($[\mathbf{x}^s \parallel X^t]$, model)
end procedure
procedure FCM–$\mathrm{U}^{\mathrm{MFC}}$ FM
    $[\mathrm{U}^{\mathrm{MFC}}, \mathbf{v}_l^s, V_l^t]$ = MFC($[\mathbf{x}^s \parallel \mathbf{Y}]$, X^t, C, $\mathbf{U}$, m, λ)
    model=trainFCM($[(\mathrm{U}^{\mathrm{MFC}})^T \parallel \mathbf{Y}]$, C, m)
    Compute distance from $\mathbf{x}^s$ to $\mathbf{v}_l^s$ and X^t to V_l^t
    Update the partition matrix $\mathrm{U}^{\mathrm{MFC}}$
    Y_m=test($\mathrm{U}^{\mathrm{MFC}}$, model)
end procedure

Since statistical methods rely on measures that consider the spreading of values and do not consider the nature of data, they often lead to the loss of important information. In the case of health care data it is important however to consider variable measurements distant from other observations, since they can represent sudden variations in a patient physiological condition. In this work, data outliers were removed using expert knowledge, meaning that values outside the acceptable physiological ranges were deleted.

4.5.2 Vasopressors Administration

MIMIC II database was used to derive models to predict the necessity of vasopressors administration in two specific subsets of patients: patients suffering from pancreatitis and patients suffering from pneumonia. Given that these patients are usually treated differently in terms of medication and surgery procedures, the circumstances related to the initiation of vasopressors are also presumably distinct; therefore models are built for each dataset separately.

The datasets contain data regarding: patients that received one of the following vasopressors: levophed, dopamine, epinephrine, vasopressin and neosynephrine; the patients' first ICU stay (in order to consider the first administration of vasopressors); a period of at least 6 h between the initiation of data acquisition and the administration of vasopressors; patients that were on vasopressors for more than 2 h.

The final pancreatitis and pneumonia datasets contain 378 and 1323 patients, respectively. Time variant features were sampled during the length of stay of the patient in the ICU, whereas time invariant features were selected from demographic information and scores records on admission. Time variant variables were selected based on a previous study (Fialho et al. 2011), where the best predictors of the need of vasopressors where determined using a combination of fuzzy modeling and bottom-up for feature selection. In order to predict in a timely manner the initiation of the vasopressor administration, a window of 2 h of data collected before the administration was not used neither for modeling nor for validation of the models. The data was then resampled considering only 10 h before the window with a sampling time of 1 h. The output consists of a binary classification with positive value if the patient was on vasopressors.

List of physiological variables, demographics and score records extracted from the MIMIC II database for vasopressors administration classification in pancreatitis and pneumonia patients:

Time variant (Pancreatitis):
- Sodium (mEq/L)
- BUN: Blood urea nitrogen (mg/dL)
- WBC: White blood cells ($\times 10^3$cells/μL)

Time variant (Pneumonia):
- Lactic acid (mg/dL)
- WBC: White blood cells ($\times 10^3$cells/μL)
- PaCO2: Arterial carbon dioxide partial pressure (mmHg)
- NBP: Non-invasive blood pressure mean (mmHg)

Time invariant (Pneumonia and Pancreatitis):
- Age (years)
- Weight (kg)
- SAPS II on admission: Simplified Acute Physiology Score
- SOFA on admission: Sequential Organ Failure Assessment

4.5.3 Mortality in Abdominal Septic Shock

The MEDAN database was used to develop classification models for mortality prediction of patients under abdominal septic shock. The pre-processing of the original data performed by Marques et al. (2011) was used, assuring data quality.

The most relevant features determined by Fialho et al. (2010) were selected, resulting in a dataset containing records of 12 time variant features measured over different periods of time, with a global sampling time of 24 h. The time series resulting from these measurements were used as the time variant input, while patients' demographic information, age and weight, formed the time invariant input. In order to maintain equal lengths of time series regarding each feature, only the last 10 days of patient care were considered, resulting in 10 sample points per feature. In this approach, patients with less than 10 measures per feature were discarded. The final dataset comprises 100 patients, from which 44 did not survive (labeled as class 1).

List of physiological variables and demographic information extracted from the MEDAN database for mortality classification in abdominal septic shock patients:

Time variant:
- Arterial pCO2 (mmHg)
- Central venous pressure (cmH2O)
- Hematocrit (%)
- Hemoglobin (g/dl)
- Heart rate (beats/min)
- WBC (cells$\times 10^3/\mu$L)
- pH
- Serum calcium (mmol/L)
- Serum creatinine (mg/dL)
- Serum sodium (mmol/L)
- Systolic blood pressure (mmHg)
- Temperature (°C)

Time invariant:
- Age (years)
- Weight (kg)

4.5.4 Readmissions

MIMIC II was used to develop models for the prediction of early readmissions. Time variant data consists of the time series representing the last 10 measurements of 7 variables, collected during the patients' stay at the ICU. The selected variables were chosen based on previous studies that made use of the same database (Fialho et al. 2012). Age, weight on admission, SAPS II and SOFA scores on admission were also collected for each patient and used as time invariant inputs. Patients

readmitted to the ICU within a period of 24–72 h after discharge and patients who only experienced one ICU stay and did not die within 1 year after discharge were respectively labeled as class 1 and class 0. The final dataset includes 2653 patients, from which 199 were readmitted within 24–72 h after discharge.
List physiological variables and demographic information extracted from the MIMIC II database for readmissions classification:

Time variant:
- Creatinine (mg/dL)
- Lactic acid (mg/dL)
- NBP: Non-invasive blood pressure mean (mmHg)
- Platelets ($\times 10^3$ cells/μL)
- Temperature (°C)
- Heart rate (beats per minute)
- SpO2: Oxygen saturation in the blood (%)

Time invariant:
- Age (years)
- Weight (kg)
- SAPS II on admission: Simplified Acute Physiology Score
- SOFA on admission: Sequential Organ Failure Assessment

4.6 Results

The performance of the models is evaluated in terms of area under the receiver operating characteristic curve (AUC) (Hanley and McNeil 1982), accuracy (correct classification rate), sensitivity (true positive classification rate) and specificity (true negative classification rate).

The datasets are evaluated using fivefold cross validation. For each fold, a grid search is performed on the training set to find the λ that maximizes the AUC; the training set is randomly divided into two smaller subsets *T1* and *T2*, where *T1* represents 50% of the training set. *T1* is used for training and *T2* for tuning. For this model, a range of values of the threshold γ are evaluated on set training set and the γ that results in the smallest difference between sensitivity and specificity is used. The model is tested using data the model has not yet used for training or tuning. For FCM–U$^{\mathrm{MFC}}$, the input U used for testing the models is obtained by computing the partition matrix using the cluster prototypes obtained in training. For updating the matrix, the distance $d^2_\lambda(\mathbf{v}^s_l, V^t_l, x_i)$ between the test entities and the cluster prototypes is computed. Cross validation is performed separately for $C = \{2, 3, \ldots, 10\}$, and results are averaged over the folds. In the end, results are shown for the C giving the best average.

Figure 4.1 shows the average performance in terms of AUC, for each C, whereas the best average results are shown in Table 4.1.

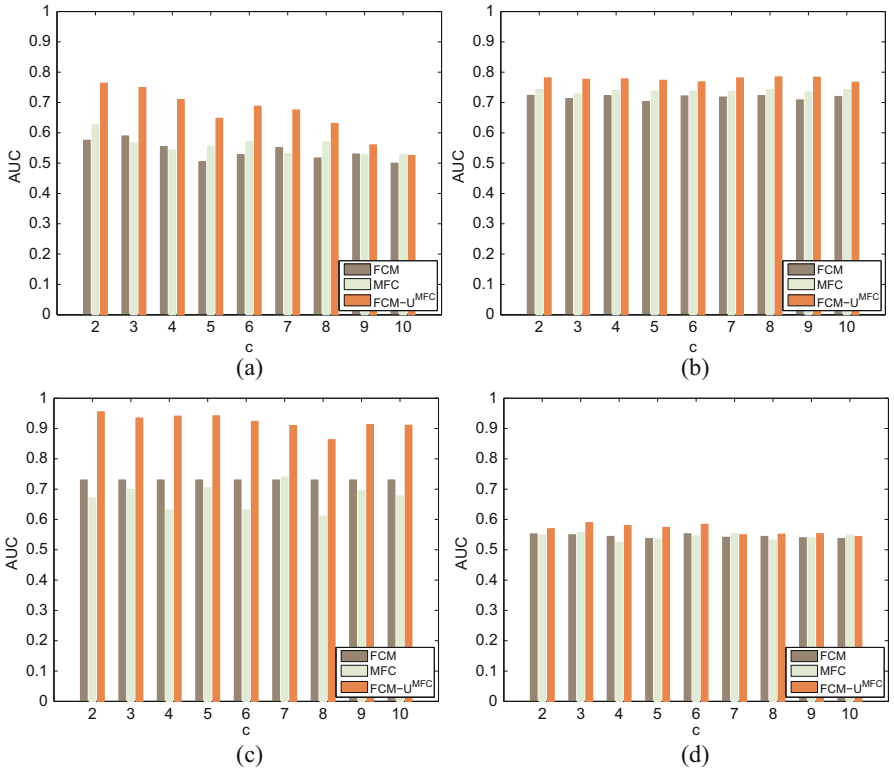

Fig. 4.1 Performance of FCM, MFC and FCM–U^{MFC} fuzzy models for different number of clusters, for MEDAN, readmissions, pancreatitis and pneumonia datasets. (**a**) Pancreatitis. (**b**) Pneumonia. (**c**) MEDAN. (**d**) Readmissions

In order to investigate the influence of λ in the performance of each method, boxplots showing the selected values of λ for different number of clusters are presented in Fig. 4.2. Overall, the choice of λ seems to be greatly affected by the data divisions, exception made to the MFC method in MEDAN and pneumonia datasets.

FCM–U^{MFC} FM performs better than the other methods in all datasets. In particular, FCM–U^{MFC} FM increases the AUC of MEDAN and pancreatitis by a factor of approximately 30%, when compared to FCM FM. MFC FM has also improved the performance of FCM FM in all datasets, due to its ability of creating rules that adapt to the mixed nature of data. Table 4.1 and Fig. 4.2 show that for all health care applications investigated, the information contained in the time variant variables is more relevant in predicting the output than the information contained in the time invariant variables.

Compared to a previous study using the same datasets (Ferreira et al. 2015), there is an overall increase in the performance of all FMs. Two main reasons can be

Table 4.1 Results with fivefold cross validation

Dataset	Fuzzy models	C	AUC	ACC	Sensitivity	Specificity
MEDAN	FCM	10	0.73 ± 0.09	0.64 ± 0.08	0.62 ± 0.14	0.66 ± 0.11
	MFC	7	0.74 ± 0.11	0.66 ± 0.09	0.61 ± 0.12	0.70 ± 0.11
	FCM–U^{MFC}	2	$\mathbf{0.96 \pm 0.05}$	$\mathbf{0.86 \pm 0.05}$	$\mathbf{0.91 \pm 0.09}$	$\mathbf{0.82 \pm 0.13}$
Readmissions	FCM	6	0.55 ± 0.03	0.53 ± 0.04	0.54 ± 0.05	0.53 ± 0.05
	MFC	3	0.56 ± 0.06	0.53 ± 0.04	0.56 ± 0.07	0.53 ± 0.03
	FCM–U^{MFC}	3	$\mathbf{0.59 \pm 0.02}$	$\mathbf{0.55 \pm 0.02}$	$\mathbf{0.58 \pm 0.02}$	$\mathbf{0.55 \pm 0.02}$
Pancreatitis	FCM	3	0.59 ± 0.08	0.57 ± 0.08	0.57 ± 0.14	0.56 ± 0.10
	MFC	2	0.63 ± 0.13	0.58 ± 0.12	$\mathbf{0.68 \pm 0.20}$	0.54 ± 0.13
	FCM–U^{MFC}	2	$\mathbf{0.76 \pm 0.05}$	$\mathbf{0.70 \pm 0.07}$	0.65 ± 0.03	$\mathbf{0.72 \pm 0.08}$
Pneumonia	FCM	2	0.72 ± 0.05	0.66 ± 0.04	0.69 ± 0.07	0.64 ± 0.04
	MFC	8	0.74 ± 0.05	0.68 ± 0.06	$\mathbf{0.72 \pm 0.05}$	0.66 ± 0.08
	FCM–U^{MFC}	8	$\mathbf{0.78 \pm 0.04}$	$\mathbf{0.70 \pm 0.02}$	0.70 ± 0.09	$\mathbf{0.70 \pm 0.04}$

attributed to this fact: first, model parameters λ and C accept a wider range of values, and second, in MIMIC II datasets, the gender is no longer used as a time invariant input variable, which means that the finding of structures based on the distribution of classes is not hampered by another binary variable.

Previous studies, namely Fialho et al. (2011, 2013), achieved notable results in the prediction of continuous vasopressors administration. However, these studies do not provide the means for predicting the initiation of therapy. Hence the comparison of performance results in Table 4.1 should not be straightforward.

Next follows a more detailed discussion of the results obtained for each dataset.

4.6.1 Vasopressors Administration

In the pancreatitis dataset, the AUC of FCM–U^{MFC} FMs decreases with increasing number of clusters (see Fig. 4.1a). Results suggest that patients in this group are better divided in two or three subgroups. In the case of FCM–U^{MFC}, λ values are close to 1, whereas for MFC, λ oscillates between 0.5 and 6, as shown in Fig. 4.2a. When the number of subgroups increases, a higher weight is in general given to the time variant part of data.

Contrarily to the previous vasopressors dataset, increasing number of clusters produce little changes in the AUC of pneumonia models (see Fig. 4.1b), showing that for this particular case, increasing the number of rules does not provide an added value in the prediction of the output. Thus, $C = 2$ should be considered for all FM approaches, with λ between 1.5 and 3.

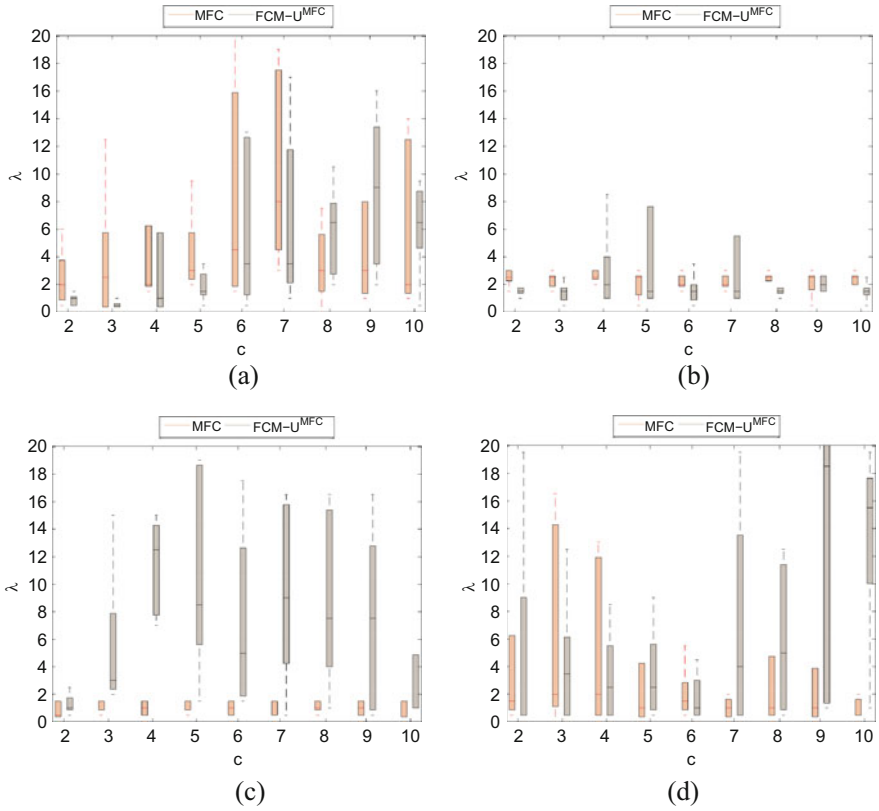

Fig. 4.2 Boxplots of λ associated to different number of clusters C, for MEDAN, readmissions, pancreatitis and pneumonia datasets. (**a**) Pancreatitis. (**b**) Pneumonia. (**c**) MEDAN. (**d**) Readmissions

4.6.2 Mortality Prediction

FCM–U$^{\text{MFC}}$ tends to select higher values of λ, except when $C = 2$, while MFC tends to select values close to 1, i.e, it gives the same weight to both time variant and time invariant components of data, as shown in Fig. 4.2c. The fact that λ is different between the rounds, oscillating between 0 and 2, justifies the differences between MFC and FCM FM approaches, highlighting the importance of this parameter in the tuning of the models. If λ would equal 1 in all rounds of MFC, results were expected to deviate less and be more similar between MFC and FCM approaches. The fact that higher values of λ are associated with improved performance also highlights the importance of this parameter.

Figure 4.1c shows that for both FCM and MFC FM, the C selected by grid search is not the best option. Smaller values of C achieve nearly the same performance, at lower computational costs and simplified model interpretability.

Thus, for this dataset, 2 clusters and 3 clusters (or rules) would be sufficient to derive FCM and MFC models, respectively. Nonetheless, the best strategy is still to perform dimensionality reduction by transforming the input variables into degrees of membership to 2 clusters.

While FCM and MFC FM approaches perform similarly to Fialho et al. (2010) when using a reduced number of features, FCM–U^{MFC} significantly improved previous results (AUC$= 0.75 \pm 0.01$ vs AUC$= 0.96 \pm 0.05$).

4.6.3 Readmissions

For this dataset, varying number of clusters result in small changes in the performance, and high values of λ are in general selected for both MFC approaches, as shown in Figs. 4.1d and 4.2d.

Monitoring signals, such as NBP mean, temperature, heart rate and SpO2 are associated with higher sampling rates than laboratory results such as lactic acid, platelets and creatinine. Having this in mind, the fact that all time series have a length of 10 points means that the data used for modeling contains laboratory measurements that can go up to 10 days of each patient' stay (when applicable, otherwise missing data is filled using the ZOH), and measurements of monitoring signals of the last 10 h of stay, approximately. We point this misalignment as the main probable reason for the overall poor results in comparison to approaches using the mean values during the last 24 h (Fialho et al. 2012). Hence, further studies should be conducted in order to handle unevenly and misaligned time series. Other reasons may be pointed out to justify the poor results, namely the highly imbalanced class distribution.

4.7 Conclusions

This work presents two modeling strategies based on the mixed fuzzy clustering algorithm, in order to handle datasets containing time variant and time invariant features, converging their information to improve knowledge extraction. One strategy uses Takagi-Sugeno where the antecedent fuzzy sets are determined by MFC in the product space of the time variant and time invariant variables and the other strategy uses Takagi-Sugeno where the antecedents are determined based on FCM in the product space of the membership degrees derived by MFC.

The performance of models is tested in four health care datasets, for the classification of critically ill patients, and is compared with Takagi-Sugeno based on FCM. The best method, FCM–U^{MFC}, is common to all applications: mortality in abdominal septic shock patients is classified with an AUC of 0.96 ± 0.05, readmissions to the ICU with 0.59 ± 0.02 and vasopressors administration in pancreatitis and in pneumonia patients with 0.76 ± 0.05 and 0.78 ± 0.04, respectively. The findings of

this work suggest that dimensionality reduction based on the transformation of input variables into degrees of membership allows the finding of important structures in data, hence the discover of relevant rules in the knowledge base system.

Future work should focus on finding which time variant and time invariant features are best predictors of different adverse events in the ICU, using wrapper feature selection methods for performance improvement.

Acknowledgements This work is funded by Portuguese Funds through the FCT, Foundation for Science and Technology under the project UID/EMS/50022/2013 and IF/00833/2014/CP1238/CT0002. S.M. Vieira acknowledges the support by Program Investigador FCT (IF/00833/2014) from FCT, co-funded by the European Social Fund (ESF) through the Operational Program Human Potential (POPH).

References

Allaudeen N, Schnipper JL, Orav EJ, Wachter RM, Vidyarthi AR (2011) Inability of providers to predict unplanned readmissions. J Gen Intern Med 26(7):771–776

Badawi O, Breslow MJ (2012) Readmissions and death after ICU discharge: development and validation of two predictive models. PloS One 7(11):e48758

Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10(2):191–203

Boudesteijn E, Arbous S, van den Berg P (2007) Predictors of intensive care unit readmission within 48 hours after discharge. Crit Care 11(Suppl 2):P475

Campbell AJ, Cook JA, Adey G, Cuthbertson BH (2008) Predicting death and readmission after intensive care discharge. Br J Anaesth 100(5):656–662

Celi LA, Mark RG, Stone DJ, Montgomery RA (2013) "big data" in the intensive care unit. closing the data loop. Am J Respir Crit Care Med 187(11):1157–1160

Clermont G, Angus DC, DiRusso SM, Griffin M, Linde-Zwirble WT (2001) Predicting hospital mortality for patients in the intensive care unit: a comparison of artificial neural networks with logistic regression models. Crit Care Med 29(2):291–296

Daly K, Beale R, Chang R (2001) Reduction in mortality after inappropriate early discharge from intensive care unit: logistic regression triage model. BMJ 322(7297):1274

Elliott M, Worrall-Carter L, Page K (2014) Intensive care readmission: a contemporary review of the literature. Intensive Crit Care Nurs 30(3):121–137

Ferreira MC, Salgado CM, Viegas JL, Schäfer H, Azevedo CS, Vieira SM, Sousa JMC (2015) Fuzzy modeling based on mixed fuzzy clustering for health care applications. In: 2015 IEEE international conference on FUZZ-IEEE

Fialho AS, Cismondi F, Vieira SM, Sousa JMC, Reti SR, Howell MD, Finkelstein SN (2010) Predicting outcomes of septic shock patients using feature selection based on soft computing techniques. In: Information processing and management of uncertainty in knowledge-based systems. Applications. Springer, Berlin, pp 65–74

Fialho AS, Cismondi F, Vieira SM, Sousa JMC, Reti SR, Celi LA, Howell MD, Finkelstein SN (2011) Fuzzy modeling to predict administration of vasopressors in intensive care unit patients. IEEE Int Conf Fuzzy Syst (ii):2296–2303

Fialho AS, Cismondi F, Vieira SM, Reti SR, Sousa JMC, Finkelstein SN (2012) Data mining using clinical physiology at discharge to predict ICU readmissions. Expert Syst Appl 39(18):13158–13165

Fialho AS, Celi LA, Cismondi F, Vieira SM, Reti SR, Sousa JMC, Finkelstein SN (2013) Disease-based modeling to predict fluid response in intensive care units. Methods Inf Med 52:494–502

Frize M, Ennett CM, Stevenson M, Trigg HC (2001) Clinical decision support systems for intensive care units: using artificial neural networks. Med Eng Phys 23(3):217–225

Frost SA, Tam V, Alexandrou E, Hunt L, Salamonson Y, Davidson PM, Parr MJ, Hillman KM (2010) Readmission to intensive care: development of a nomogram for individualising risk. Crit Care Resusc 12(2):83–89

Gajic O, Malinchoc M, Comfere TB, Harris MR, Achouiti A, Yilmaz M, Schultz MJ, Hubmayr RD, Afessa B, Farmer JC (2008) The stability and workload index for transfer score predicts unplanned intensive care unit patient readmission: Initial development and validation*. Crit Care Med 36(3):676–682

Goldfield NI, McCullough EC, Hughes JS, Tang AM, Eastman B, Rawlins LK, Averill RF (2008) Identifying potentially preventable readmissions. Health Care Financ Rev 30(1):75

Hanisch E, Brause R, Arlt B, Paetz J, Holzer K (2003) The MEDAN Database. http://www.medan.de.

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(4):29–36

Hug CW, Szolovits P (2009) ICU acuity: real-time models versus daily models. In: Association AMI (ed) AMIA annual symposium proceedings, vol 2009, p 260

Izakian H, Pedrycz W, Jamal I (2013) Clustering spatiotemporal data: an augmented fuzzy C-means. IEEE Trans Fuzzy Syst 21(5):855–868

Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, Kripalani S (2011) Risk prediction models for hospital readmission: a systematic review. J Am Med Assoc 306(15):1688–1698

Knaus WA, Draper EA, Wagner DP, Zimmerman JE (1985) Apache II: a severity of disease classification system. Crit Care Med 13(10):818–829

Le Gall JR, Lemeshow S, Saulnier F (1993) A new simplified acute physiology score (saps II) based on a European/North American multicenter study. J Am Med Assoc 270(24):2957–2963

Marques FJ, Moutinho A, Vieira SM, Sousa JMC (2011) Preprocessing of clinical databases to improve classification accuracy of patient diagnosis. IFAC Proc Vol (IFAC-PapersOnline) 18:14121–14126

Ouanes I, Schwebel C, Français A, Bruel C, Philippart F, Vesin A, Soufir L, Adrie C, Garrouste-Orgeas M, Timsit J, Misset B (2012) A model to predict short-term death or readmission after intensive care unit discharge. J Crit Care 27(4):422–e1

Paetz J (2003) Knowledge-based approach to septic shock patient data using a neural network with trapezoidal activation functions. Artif Intell Med 28(2):207–230

Paetz J, Arlt B, Erz K, Holzer K, Brause R, Hanisch E (2004) Data quality aspects of a database for abdominal septic shock patients. Comput Methods Prog Biomed 75:23–30

Reini K, Fredrikson M, Oscarsson A (2012) The prognostic value of the modified early warning score in critically ill patients: a prospective, observational study. Eur J Anaesthesiol (EJA) 29(3):152–157

Rosenberg AL, Watts C (2000) Patients readmitted to ICUs*: a systematic review of risk factors and outcomes. CHEST J 118(2):492–502

Rosenberg AL, Hofer TP, Hayward RA, Strachan C, Watts CM (2001) Who bounces back? Physiologic and other predictors of intensive care unit readmission. Crit Care Med 29(3):511–518

Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L, Moody G, Heldt T, Kyaw TH, Moody B, Mark RG (2011) Multiparameter intelligent monitoring in intensive care II (MIMIC-II): a public-access intensive care unit database. Crit Care Med 39(5):952

Salgado CM, Vieira SM, Mendonça LF, Finkelstein S, Sousa JMC (2016) Ensemble fuzzy models in personalized medicine: application to vasopressors administration. Eng Appl Artif Intell 49:141–148

Strand K, Flaatten H (2008) Severity scoring in the ICU: a review. Acta Anaesthesiol Scand 52(4):467–78

Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modeling and control. IEEE Trans Syst Man Cybern 15(1):116–132

Vieira SM, Carvalho JP, Fialho AS, Reti SR, Finkelstein SN, Sousa JMC (2013) A decision support system for ICU readmissions prevention. In: IFSA world congress and NAFIPS annual meeting (IFSA/NAFIPS), 2013 joint. IEEE, Piscataway, NJ, pp 251–256

Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG (1996) The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. Intensive Care Med 22(7):707–710

Walsh C, Hripcsak G (2014) The effects of data sources, cohort selection, and outcome definition on a predictive model of risk of thirty-day hospital readmissions. J Biomed Inform 52:418–426

Zheng B, Zhang J, Yoon SW, Lam SS, Khasawneh M, Poranki S (2015) Predictive modeling of hospital readmissions using metaheuristics and data mining. Expert Syst Appl 42(20):7110–7120

Chapter 5
Operations Research for Occupancy Modeling at Hospital Wards and Its Integration into Practice

N.M. van de Vrugt, A.J. Schneider, M.E. Zonderland, D.A. Stanford, and R.J. Boucherie

5.1 Introduction

Medical and technological advancements are contributing to increase healthcare expenditures and increase numbers of hospitalized patients (Chernew and Newhouse 2012), while at the same time the length of stay (LoS) for these patients decreases. However, healthcare expenditures are still rising (OECD). Society calls for improved cost effective healthcare delivery, which puts pressure on available

N.M. van de Vrugt (✉)
Centre of Operations Improvement and Research (CHOIR) and Department of Applied Mathematics, University of Twente, Postbox 217, 7500AE Enschede, The Netherlands

Jeroen Bosch Hospital, Henri Dunantstraat 1, 5223 GZ's Hertogenbosch, The Netherlands

Leiden University Medical Center, Leiden, The Netherlands
e-mail: n.m.vandevrugt@utwente.nl

A.J. Schneider
Centre of Operations Improvement and Research (CHOIR) and Department of Applied Mathematics, University of Twente, Postbox 217, 7500AE Enschede, The Netherlands

Leiden University Medical Center, Quality of Care-Institute, Albinusdreef 2, 2333 ZA Leiden, The Netherlands
e-mail: a.j.schneider@lumc.nl

M.E. Zonderland • R.J. Boucherie
Centre of Operations Improvement and Research (CHOIR) and Department of Applied Mathematics, University of Twente, Postbox 217, 7500AE Enschede, The Netherlands

D.A. Stanford
Department of Statistical and Actuarial Sciences, University of Western Ontario, 1151 Richmond Street North, London, ON, Canada N6A 5B7

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_5

financial resources. At present, hospitals tend to focus on process improvement by decreasing unnatural (e.g. self-created) variability and alignment of different services.

During hospitalization, patients spend most of their time in wards. These wards are also referred to as inpatient care facilities, and provide care to hospitalized patients by offering a room, a bed and board. Wards are strongly interrelated with upstream hospital services such as the operating theater and the emergency department. Due to this interrelation it is essential to attain a high efficiency level at hospital wards in order to achieve efficient patient flow. Hospital ward management often aims for bed occupancy rates above 85% in order to maximize throughput, leaving little slack for flow fluctuations which results in refused, deferred and/or rescheduled patients.

Operations Research (OR) can give managerial insights about trade-offs between performance indicators, such as bed occupancy rates and blocked patients. Although OR methods have the potential to lead to large improvements in all sorts of processes, it appears that the cases in which the models and/or results have been actually implemented are sparse. Using OR models, possible interventions can be evaluated in a safe environment, reducing the risk of implementing an intervention that appears to be counter-productive.

OR models may be invoked for different objectives, for example to provide insights or to optimize a certain performance measure and what effects changes in demand or supply have on these performance indicators. The logistically important performance measures for hospital wards are throughput, blocking probability and occupancy. A possible objective in this area could be to determine the optimal capacity, warranting a prespecified maximal blocking probability and minimal occupancy levels.

In this chapter we focus on occupancy modeling. Related topics, not covered in this chapter, are for example optimizing the assignment of patients to beds, and optimizing patients' access times. The bed assignment problem becomes important when, for example, a ward accommodates patients with infectious diseases, or patients that do not share rooms with the opposite sex. A patient's access time is the number of days between the request for an appointment and the appointment itself, and may be improved by optimizing patient admission schedules and/or the operating theater schedule. Additionally, material logistics and facility design problems are outside the scope of this chapter.

Our aim in this review chapter is to guide both researchers and healthcare professionals through the OR concepts which have been applied to hospital wards. We first present the terminology on the different types of wards and performance indicators covered in this chapter. Next, we give an overview of articles where OR techniques are applied to ward related problems, followed by some detailed examples on how to apply these models. We conclude the chapter by looking at the integration of OR models into practice and possibilities for further research. As background information, we provide a brief introduction of OR models in the appendix.

5.2 Hospital Ward Types and Terminology

In this section we introduce the types of hospital wards and performance measures as used in this chapter. Throughout this chapter, we define a hospital ward as follows: an area or unit within a hospital where inpatients with comparable medical conditions are admitted to a bed to receive care. This typically involves staying overnight until their medical condition changes in such a way that the patient either leaves the hospital, or is transferred to a ward with a different level of care. We therefore place the beds associated with the operating theater (OT), the emergency department (ED) or the outpatient clinics outside the scope of this chapter, as they usually temporarily accommodate patients that undergo a (short) treatment.

The logistical performance of wards is generally assessed by three indicators which are related to each other: throughput, blocking probability and occupancy. The exact definitions of these three performance indicators is given in Sect. 5.2.2, after our definitions of different ward types.

5.2.1 Taxonomy

In this section we distinguish different ward types based on logistical characteristics: the type of in- and outflow, typical length of stay (LoS) and resources, and planning problems the wards face. Based on the literature cited in this chapter, we distinguish the following types of wards:

- Intensive Care Unit (ICU)
- Acute Medical Unit (AMU)
- Obstetric ward (OBS)
- Weekday ward (WDW)
- General ward

We describe each type of ward in terms of (logistical) characteristics below and demonstrate why it is a different type of ward. An overview of the differences is summarized in Table 5.1, in which '0' denotes average occurrence, occupancy or costs, and '+' ('−') denotes increased (decreased) compared to average.

Intensive Care Unit (ICU) For this category in our taxonomy we group several ward types with similar logistical characteristics: traditional ICUs, specialized ICUs, and Critical, High or Medium Care Units. Specialized ICUs are, for example, stroke units, cardiac care units and neonatal ICUs. High Care and Medium Care Units are sometimes combined and often referred as step-down units between ICUs and general wards. In the United Kingdom these combined wards are also referred 'Critical Care Units'. The difference between high and medium care is generally the necessity of breathing support. The ICU of a hospital accommodates the most severely ill patients who require constant close monitoring and support from advanced medical equipment and staff (nurses mostly on a 1:1 basis and intensivists which are readily available) (Mallor and Azcarate 2014). In the remainder of this chapter we refer to the ward types discussed in this section as 'ICUs'.

Table 5.1 Summary of characteristics per ward type

	ICU	AMU	WDW	OBS	General
Long LoS	+	−	−	0	+
Short LoS	+	+	+	+	0
Acute admissions	+	+	−	+	+
Elective admissions	+	−	+	+	+
Bed occupancy	+	−	+	−	0
Staff/bed ratio	+	+	−	0	0
Equipment	+	0	−	0	0

Due to the used equipment and available staff the ICU has the highest costs per bed of all hospital wards. An ICU preferably doesn't defer patients, as this would imply serious mortality risks. However, the costs per bed do not allow for a large buffer in the number of available beds. Therefore, ICUs tend to be fully occupied, and discharge the least ill patient when a bed needs to be freed for a newly arriving patient, or cancel an elective procedure at the OT which requires ICU capacity afterwards. Patient typically either have a short LoS or a very long LoS, and arrive from the OT, ED, wards or surrounding hospitals.

Acute Medical Unit (AMU) AMUs lack a uniform definition. We think the following definition covers the best definition of AMUs 'an AMU is a designated hospital ward specifically staffed and equipped to receive medical inpatients presenting with acute medical illness from EDs and outpatient clinics for expedited multidisciplinary and medical specialist assessment, care and treatment for up to a designated period (typically between 24 and 72 h) prior to discharge or transfer to medical wards' (Scott et al. 2009). Often, AMUs serve as a buffer for both the ED and inpatient wards. Since an AMU treats only urgent patients and should alleviate ED congestion, management is more focused on throughput and LoS, and the target utilization of the AMU beds is typically lower compared to general wards. AMUs are also known under synonyms as 'emergency observation and assessment ward, 'acute assessment unit' and 'acute medical assessment units'. The review papers available (Cooke et al. 2003; Scott et al. 2009) provide a comprehensive overview of definitions and concepts for AMUs. The inflow mainly consists of acute patients from the ED, outpatient clinics, surrounding hospitals or General Practitioners.

Weekday Ward (WDW) WDWs are wards admitting patients with an expected LoS between 2 and 5 days, which are usually only open on weekdays (Conforti et al. 2011). WDW-type of hospitals are also sometimes referred to as 'Monday to Friday clinic' or 'Week Hospital'. Most patients at WDWs are elective, and can be transferred to regular wards without any health risks. Only patients with a highly predictable LoS may be admitted, which is why WDWs mostly treat patients for which strict treatment protocols apply. Scheduling patients at a WDW is complicated by each patient's different LoS and urgency level, which implies a deadline by which the patient should be treated. The requirement that the ward should be closed during weekends also complicates patient scheduling. Most admissions arrive directly from home.

Obstetric Ward (OBS) Obstetric and Gynecology wards provide care for women during their pregnancy, during and after labor, and also take care of their newborns (Cochran and Bharti 2006). Additionally, Gynecology wards accommodate women with problems regarding their reproductive organs. The women at these wards often require (brief) surgical intervention, and typically a short hospitalization. Some hospitals group these types of wards under names like 'Birthing Center', 'Maternity Clinic', or 'Women's and Child's Center'. Most patients arrive from home, outpatient clinics or other hospitals.

General Wards General wards in hospitals are often dedicated to a single medical specialty such as Neurology, Geriatrics, or Hematology. As these wards are generally equipped with similar resources and accommodate both acute and elective admissions which and differ in LoS, we aggregate these ward types. General wards can either be surgical or medical and some wards, such as psychiatric or geriatric wards, are closed, implying that patients cannot leave the wards without approval. Other wards are equipped with a specific type of resource, such as dialysis machines and heart monitors. The nurse to patient ratio is often 1:5–1:6. Patients with a particular medical specialty are typically not all accommodated in the same ward, but may also be admitted at for example a WDW or an ICU. Patient inflow is mainly formed by referrals of outpatient clinics, ICUs, General Practitioners or other hospitals.

5.2.2 Terminology

In healthcare a concept such as 'occupancy', which may seem simple at first sight, has several different definitions. Different researchers and healthcare practitioners use different definitions of occupancy, which may result in false comparisons when the used definitions are not clearly stated. Therefore, we define the frequently used concepts in the following paragraphs. We first define different concepts of capacity (based on Vissers and Beech 2005), then define the throughput and blocking probability, and conclude this section with the different concepts of occupancy.

Each ward has a certain capacity, which is expressed in terms of the number of patients and their care intensity that the ward can accommodate. The capacity of a ward is measured by the number of beds and nurses, and there are different types of capacity. The physical capacity is the number of beds at the ward. Each nurse can take care of a certain number of patients in parallel (determined by the nurse to patient ratio), which determines the structural available capacity. Additionally, temporary capacity changes can occur; for example bed closures in holiday periods, or beds that are used which are officially not staffed in case of bed shortage. The structural capacity and temporary changes together determine the (average) realized available capacity.

Suppose, in a highly stylized example, that a hospital ward has 15 beds in a certain area. There are always three nurses scheduled to work at the ward, and each nurse can take care of at most four patients at the same time. Each summer and Christmas holidays the ward experiences decreasing patient numbers, and decides to only schedule two nurses. The holiday periods together last 8 weeks. Then, for this ward the physical capacity is 15 beds, and the structural capacity is 3 (nurses) $\times$4 (patients per nurse) $=$ 12 beds. Due to the holidays, each year has 8 weeks in which only eight beds are open, so the average realized capacity is:

$$\frac{8(\text{weeks}) \times 8(\text{beds}) + (52-8)(\text{weeks}) \times 12(\text{beds})}{52(\text{weeks})} \approx 11.4 \text{ beds}.$$

As mentioned in the introduction of this section, the logistical performance of a ward is assessed by three performance indicators: throughput, blocking probability and occupancy. These indicators are all related to each other. The throughput of a ward can be measured as the number of admissions or discharges per time unit. The blocking probability of a ward is the percentage of patients that request a bed at the ward at an instance that there are no available beds:

$$P_b = \frac{\text{Number of patients not accommodated at ward}}{\text{Total number of patients requesting a bed at ward}} \times 100\%. \tag{5.1}$$

Blocked patients are either accommodated in a different ward, or deferred to another hospital.

In contrast to throughput and blocking probability, bed occupancy can be quantified by three definitions: based a on bed census at certain time, based on real LoS or based on the number of hospitalization days. Here we aim to give an overview of the most commonly used definitions.

One of the definitions of bed occupancy includes the bed census measured once a day at a specified point in time, for example every morning at 10:00 am. Then, dividing the average of these measurements by the structural available capacity, the occupancy is:

$$O_{bc}(t) = \frac{\text{average bed census at time } t}{\text{structural available capacity}} \times 100\%. \tag{5.2}$$

Note that for the occupancy it also matters how the capacity of a ward is calculated; in most hospitals the structural available capacity is used. A slightly different occupancy measure is obtained by taking the average of multiple bed census measurements throughout each day, for example each hour; we denote this measure by $\bar{O}_{bc}$. The advantage of taking more measurements is that it will better reflect actual bed usage.

Hospitals may also define the occupancy of a ward as the ratio between the total time patients were in beds at the ward and the total time available:

$$O_{LoS}(T) = \frac{\text{sum of all LoSs for all patients in time period } T}{\text{structural available capacity} \times \text{time period } T} \times 100\%. \quad (5.3)$$

This measure is calculated using admittance and discharge time stamps for a certain measurement period, or by multiplying the average LoS with the number of patients accommodated at the ward. This occupancy measure reflects the actual time the beds are used, but does not incorporate unavailability due to cleaning of beds.

Until recently, it was common in Dutch hospitals to determine the bed occupancy using the hospitalization days declared to the insurance companies:

$$O_{hd}(T) = \frac{\text{sum of hospitalization days for all patients in } T}{\text{structural available capacity} \times \text{length } T} \times 100\%. \quad (5.4)$$

Financial hospitalization days were counted in integers, and could be declared if the patient is in a bed before 8:00 pm and discharged after 7:00 am the next day. This implied that the occupancy could be over 100% as beds can be reused if patients are discharged early in the day and new patients are admitted in the afternoon. A drawback of this measure is that it cannot be used as a targeted occupancy for all ward types. Such a situation would arise in wards in which patients generally stay for only a part of a day so that multiple patients can be served by the same bed on the same day (e.g. gynecology). In this system, these wards should therefore achieve occupancy targets over 100%, while wards at which patients stay much longer (e.g. geriatrics) will suffer severe bed shortages if the occupancy is over 90%.

This is an example of an arrival and discharge process at a ward, in order to illustrate the different concepts of occupancy. Consider a ward with three beds that is empty at the start of our observation period. We choose to observe the ward from 8:00 am on day 1, until 5:00 pm on day 4. In this period the following patients arrive:

	Arrival		Discharge			
	Day	Time	Day	Time	LoS	Hosp. days
Patient 1	1	8:00 am	2	6:00 pm	1.42	2
Patient 2	1	10:00 am	4	8:00 am	2.92	4
Patient 3	1	3:00 pm	2	8:00 am	0.71	2
Patient 4	2	3:00 am	Patient is blocked		–	–
Patient 5	2	9:00 am	3	8:00 am	0.96	2
Patient 6	3	9:00 am	After day 4		1.33	2
Patient 7	4	10:00 am	After day 4		0.29	1

(continued)

In this example, patient 4 is blocked as patients 1, 2 and 3 fill up all available beds and the first patient that is discharged (patient 3) is not discharged before patient 4 arrives. Note that the LoS for patients 6 and 7 in the table is not their exact LoS but only the part until the end of the observation period. The bed census for this ward is depicted in Fig. 5.1. The blocking probability for this time period equals 1/7≈15%. The different occupancy measures are calculated as follows.

The bed census at 10:00 am for day 1 to 4 is 2, 3, 2, and 1, respectively, so the average equals 2. Therefore $O_{bc}(10\text{ am}) = 2/3 \approx 66.7\%$. The average hourly bed census is 2.2, so $\bar{O}_{bc} = 2.2/3 \approx 74.8\%$.

The sum of the LoS for all patients at this ward in this observation period, T, equals 7.63 days. The length of the observation period is 3.38 days. Therefore, $O_{LoS}(T) = 7.63/(3 \times 3.38) \approx 75.3\%$.

The sum of the hospitalization days declared for these patients is 13, and the total number of days in this observation period is four. Therefore, $O_{hd}(T) = 13/(3 \times 4) \approx 108.3\%$.

The occupancy measure with hospitalization days is always higher than the other occupancy measures. The ordering of the remaining concepts of occupancy depends on the ward studied.

Hospital management determines which of the aforementioned occupancy measures is used, and sets the target throughput level for each ward separately. A high occupancy usually results in a high blocking probability (Bailey 1952). Therefore it is important for management to balance these three performance indicators. Adequate targets for the performance indicators depend on many factors, for example: the capacity of a ward, the fraction of admissions that is acute, the

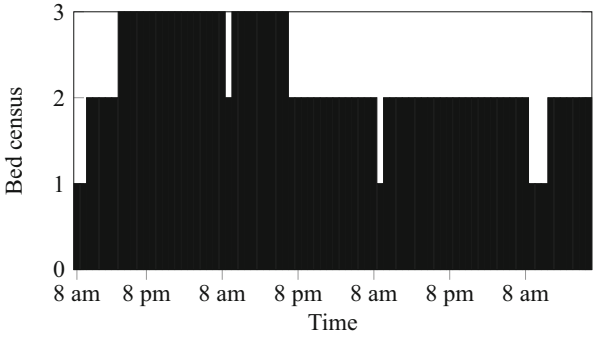

Fig. 5.1 Bed census for example

possibility of deferring admissions, the cost per bed, and the ward layout. Large wards have economies of scale, so a higher bed occupancy can be achieved with a lower blocking probability. If a ward has mostly acute admissions, occupancy targets need to be set lower; elective admissions can be rescheduled in case of bed shortage, while acute admissions cannot. If the deferral of an arriving patient could give rise to life threatening situations (e.g. in case of an intensive care unit), a ward has to lower the target occupancy to produce a lower blocking probability. However, such wards usually have high costs per staffed bed, driving the occupancy targets upwards. Finally, if a ward has many rooms with multiple beds, the bed assignment is less flexible compared to wards with many single bed rooms; if, for example, a patient has an infectious disease he cannot share a room with others. Concluding, it can be said that determining adequate occupancy, blocking probability and throughput targets is a challenging task.

5.3 Ward-Related OR Models

In the previous sections we distinguished different ward types and their logistical similarities and differences. In this section we will review the OR literature for each ward type, emphasizing the main questions or problems the literature tries to solve, the context of problems (e.g. ward type) and the type of models invoked for each paper. An overview and brief explanation of each OR method is given in the Appendix. In Table 5.2 the number of papers found for each ward type and OR model/method is displayed. If a paper invoked multiple OR models, we categorized this paper in all applicable categories. For each ward type, we review the related literature in the following subsections.

Table 5.2 Literature categorized by applied models and ward type

OR model/method	IC	AMU	WDW	OBS	General	**Total**
Algorithms	1	0	0	0	3	**4**
Dynamic Programming	1	0	1	0	0	**2**
Markov processes	4	0	0	2	11	**17**
Mathematical programming	4	2	1	3	6	**16**
Queueing theory	15	2	0	3	16	**36**
Regression	1	0	0	0	1	**2**
Simulation	22	1	0	2	21	**46**
Stochastic models	1	1	0	0	3	**5**
Time series	1	0	0	0	2	**3**
Total	**50**	**6**	**2**	**10**	**63**	**131**

5.3.1 Intensive Care Unit

At the ICU both elective and emergency patients arrive. Emergency patients mostly come from the ED or surrounding hospitals and elective patients mainly arrive after surgery. Since significant costs are involved, management tends to maximize utilization. This results in an increasing number of refusals and/or severe ill patients being transferred from the ICU to high, medium or regular care wards which could lead to situations where quality of care is at stake, and possibly to disruptions in the operating theater schedule. These are also the main problems the literature of this section focuses on: admission and discharge control. See Table 5.3 for an overview of the cited literature in this section.

A queueing model ($M/G/s/s$ queue, see the Appendix for explanation) was used to analyze the total minimum of required ICU beds for burn care for the state of New York (Blair and Lawrence 1981). The authors start by finding the number of beds at an aggregate level given a maximum blocking probability of 5%, and then apply a heuristic allocating these ICU beds among several regional units, while trying to maintain the same blocking probability. This model is extended to analyze an overflow model (Litvak et al. 2008). Here each ICU reserves bed capacity for regional emergency patients, which may be used as overflow beds in a certain region. To approximate the blocking probability of this overflow model the Equivalent Random Method is used, while a simulation model is used to validate the results of this queueing model with historical data. Another modified $M/M/s/s$ model is used to analyze different admission policies and their relation to survival gains (Shmueli et al. 2003). The policies consisted of: (1) the standard first come first served (FCFS) discipline; (2) arrivals are served if and only if a bed is available and the survival gain is greater than an arbitrary threshold value; and (3) arrivals are served if and only if a bed is available and the survival gain threshold value is met, where in this policy the threshold value is depending on the number of beds available. (If fewer beds are available, the threshold value for survival gain will increase.) The results show significant increase in survival gain in both the second and third policy compared to the first policy. The third policy showed only marginal survival gain compared to the second policy, while the number of rejected patients increased significantly. Another application of the $M/M/s/s$ queue is used for analyzing an ICU (McManus et al. 2004). This model is validated with observed data and it is proved that the calculated blocking probabilities from the queueing model were accurate.

Next to queueing, discrete time Markov chains are also applied to ICUs. The authors developed a Markov chain in order to analyze so called bumping (patient transfers from the ICU to free capacity for new arrivals which are more severely ill) (Dobson et al. 2010). Another application is used for the effect of ICU discharge strategies and bed census on patient mortality and total readmission load (patients that are hospitalized shortly after their last admission for the same medical condition) (Chan et al. 2012).

Table 5.3 Literature on ICUs categorized by applied models

OR model/method	References
Algorithms	Blair and Lawrence (1981)
Dynamic programming	Chan et al. (2012)
Markov processes	Blair and Lawrence (1981), Broyles et al. (2010), Dobson et al. (2010), Garg et al. (2010)
Mathematical programming	Kokangul (2008), Mallor and Azcarate (2014), Mallor et al. (2016)
Queueing theory	Blair and Lawrence (1981), de Bruin et al. (2007); De Bruin et al. (2010), Griffiths et al. (2013a), Kim et al. (1999), Litvak et al. (2008), Mallor et al. (2016), McManus et al. (2004), Shmueli et al. (2003), van Dijk and Kortbeek (2009), Williams et al. (2015), Yang et al. (2013), Zonderland and Boucherie (2012)
Regression	Mallor and Azcarate (2014)
Simulation	Bountourelis et al. (2011, 2013), Costa et al. (2003), Davies (1994), Kim et al. (1999), Kim et al. (2000), Kokangul (2008), Kolker (2013), Litvak et al. (2008), Mallor and Azcarate (2014), Mallor et al. (2016), Marcon et al. (2003), Marmor et al. (2013), Masterson et al. (2004), Mustafee et al. (2012), Nguyen et al. (2003), Ridge et al. (1998), Shahani et al. (2008), Sissouras and Moores (1976), Troy and Rosenberg (2009), Yang et al. (2013)
Time series	Garg et al. (2010)

Simulation is also often applied to analyze the required number of ICU beds. In Ridge et al. (1998), Kokangul (2008) and Marmor et al. (2013) they analyze several scenarios, for instance reserving ICU beds for emergency arrivals using simulation. Kim et al. (1999) simulate several ICU arrival processes and compare these results with theoretical results using an $M/M/s$ queue. Based on the simulation model, the authors also determine the blocking probability for the current capacity. Another study (Kim et al. 2000) analyzes several scenarios to minimize the number of elective surgery patients refused at the ICU. The efficient frontier method is used to plot the trade-off between the number of canceled surgeries and the average waiting time per scenario.

Some studies combine several OR techniques to analyze the ICU (Mallor and Azcarate 2014): first, a regression model is proposed for modeling the ICU LoS; second, a comprehensive simulation model is developed for analyzing system behavior and blocking probabilities; and last mathematical programming is used to model the triage problem (which current and arriving patients require ICU capacity the most?) for early or delayed discharges from the ICU depending on high or low utilization of ICU capacity.

When analyzing patient logistics at the ICU, there is a clear distinction between the type of models used and the type of problems solved. Because a significant part of the arrivals at the ICU is unscheduled, queueing theory gives accurate and

representative results. To analyze ICU dynamics, this technique is typically used to achieve general insights about blocking probability, occupancy, ICU capacity, and their mutual trade-offs. Markov chains are used to analyze bed census probabilities and the probability of bumping. Simulation is generally used to analyze multiple scenarios where particular details are involved and/or case-specific dynamics need to be studied.

5.3.2 Acute Medical Unit

The reviews on AMUs mentioned in Sect. 5.2.1 conclude that AMUs may have many advantages, but also that the evidence of economic effectiveness is thin. The AMU 'performance is dependent on good management and availability of diagnostic services', and asserted that there is no proof of cost-effectiveness of AMUs (Cooke et al. 2003). An extensive list of success factors for AMUs is also available (Scott et al. 2009). From an OR perspective, if a hospital does not add beds or staff to its current capacity for opening an AMU, the improved performance reported in the reviews is disputable. The beds assigned to the AMU are taken from other wards, decreasing the benefits of economies of scale and affecting other patients at those wards, and additionally, patients that require inpatient care after their stay at the AMU encounter more process steps than if they would have been admitted directly. Therefore, the effects of opening an AMU cannot be predicted beforehand without the use of appropriate mathematical models. Perhaps partly since AMUs are a relatively new concept, the OR literature with an AMU application is somewhat sparse. In this section, we review this available literature.

Depending on the performance measures of interest and research goals, several models could be applied to AMUs. We describe a goal programming approach used to minimize the delay from ED to AMU, and two different queueing networks to evaluate blocking probability and bed census.

A goal programming approach to determine the required additional resources (beds, doctors and nurses) for each hour of the day to minimize the delay patients experience on an AMU staffed with eight beds, two nurses and three doctors is used (Oddoye et al. 2007). Goal programming is an extension to mathematical programming, in which for each, typically conflicting, objective a target (or goal) is set and deviations from these targets are minimized. In the model, each patient requires a bed, and a specific treatment by a nurse, doctor, or both. A patient is delayed if there are no beds available upon arrival, or if the doctors and nurses are seeing other patients at the moment the patient requires care. For the case studied, the average LoS is 5 h, and the run time of the model equals a day and a half. The conclusion is that only two doctors are required, and a third nurse should be standby in the afternoon and at midnight to cope with peak demand.

In a follow-up study for a larger AMU (currently 58 beds), a simulation study analyzes 14 scenarios with different numbers of beds (Oddoye et al. 2009). Here, each resource type (beds, nurses, and doctors) has its own queue, and patients wait

in these queues until the resource they require is available. Initial targets for each queue length are fed into a goal programming model, together with targets for total LoS and the number of beds. The authors minimize weighted positive deviations from these targets. The model output comprises the resource levels that minimize patients' delay at the AMU, and a trade-off between economic objectives, i.e. higher utilization of resources, and patient- and staff-related objectives is provided.

Another study analyzes a network with one AMU and an aggregated regular ward, in which patients are transferred between the wards if their care requirements change (Utley et al. 2003b). The authors use an infinite server queueing network to determine the probability that the bed occupancy on either ward exceeds a certain number of beds. Based on this probability, they determine the optimal assignment of the available beds to either the AMU or the regular ward. In case the total mean bed occupancy is 85%, and 91% of the patients require acute care, they conclude that 60–65% of the available beds should be designated for acute care.

For a network comprising an ED, two aggregated wards, and an AMU, one study determines the blocking probability by invoking a network of Erlang loss queues in which the AMU both has direct patient arrivals and serves as an overflow ward (Zonderland et al. 2015). They consider both urgent patients (arriving from the ED) and elective patients. The hospital is only allowed to reallocate existing beds from the wards to the AMU. The equivalent random method is used to analyze the network with overflows, since overflow traffic does not follow a Poisson distribution. This method approximates the original network by truncating an infinite server network. The authors conclude that opening an AMU is beneficial for accommodating urgent patients, but the blocking probability for elective patients increases significantly.

The advantage of a simulation or goal programming approach over queueing networks, is that time-dependent arrivals can be incorporated relatively easy. However, the size of the state space in a goal programming model increases with the time horizon considered, and will explode when several departments of realistic sizes are considered. The drawback of simulation models is that they are not easily applied to other hospitals. The advantage of considering infinite server queues is that straightforward formulas for the analysis exist in the literature.

5.3.3 Obstetrics Ward

There are several OR models that have been applied to OBS wards and maternity clinics in the literature. We describe different queueing theory approaches, a simulation model, a discrete time conditional phase type model, and a discrete time Markov model.

In research conducted almost 40 years ago, the bed occupancy at an OBS ward using an infinite server queue is modelled (McClain 1978). The ward may also admit gynecology patients to achieve higher occupancy rates, but those patients are transferred to other wards if an OBS patient has no available bed upon arrival. The

gynecology patients may only be admitted to the OBS ward when the bed census is lower than a certain threshold. They use an infinite server queue to represent the situation where patients are placed in unstaffed beds as a temporary measure when no official beds are available upon arrival. The results are compared for multiple hospitals when including the national guidelines regarding the admittance of gynecology patients to OBS wards, and state which thresholds are best for certain ward sizes.

Another study calculates the probability of delay, e.g. the probability that there is no bed available upon arrival, using an $M/M/s$ queue (Green and Nguyen 2001). Key to this model is that arriving patients who find all beds occupied wait at the clinic until a bed becomes available. During their waiting time, patients are not treated, as their 'service' commences as soon they are placed in a bed. Inputs are the average LoS found in hospital data and different arrival rates. The authors compare the probability of delay for different occupancy targets and different arrival rates.

For a maternity clinic consisting of different wards, including a neonatal ward and ICU, the Queueing Network Analyzer is used (cf. Zonderland and Boucherie 2012, Sect. 2.4.4) to model the bed occupancy (Cochran and Bharti 2006). The authors evaluate all possible bed arrangements among the wards for the peak arrival rate of the clinic. The best arrangements are then evaluated in a system with an inhomogeneous arrival rate in a Discrete Event Simulation. The authors report that the hospital has implemented some of their recommendations, but instead of reassigning beds the hospital chose to add 15 beds to the ward with the highest bed shortage according to the simulation and the queueing model.

To model different types of wards in a network of multiple maternity clinics independent $M/M/s/s$ queues are also used (Pehlivan et al. 2012). The general Erlang loss formulas for the blocking probability are then fed into a Mixed Integer Linear Program to determine strategic bed assignment policies. Each year the clinic may reassign, open and close beds at the wards and clinics, and each decision entails certain costs. The authors incorporate long term planning, since it is undesirable that one year, a ward closes beds and fires nursing staff, while the next year, these beds are reopened and staff are recalled. The objective of the optimization program is to minimize the costs over the decision horizon. One of their conclusions is that efficiency could be gained if resources are transferred among units that experience different demographic changes (increase or decrease in the number of women giving birth).

In an attempt to improve the occupancy rate of an obstetric clinic, one study investigates different scenarios by means of Discrete Event Simulation (Griffin et al. 2012). Inflow and LoS of the model are based on hospital data; patients in the model follow one of the predefined care pathways through the clinic. The authors conclude that the care pathway based approach reflects reality better than a transition probability based approach when they compare the results of both approaches to hospital data. One of the investigated scenarios includes 'swing rooms', which are rooms that can be used by multiple wards of the clinic, but not at the same time. The

clinic implemented the swing rooms, which proved useful for balancing utilization throughout the clinic during bed census peaks.

A discrete time Markov model is developed to mimic a maternity clinic consisting of four wards (Isken et al. 2011). Patients can flow among units, with the routes patients take depending on their type. The authors define eleven patient types and six arrival streams (e.g. natural birth or cesarean), and the LoS has an empirical discrete distribution. All input is derived from hospital data. Since the model assumes infinite capacity, the authors derive the mean and variance of the bed occupancy at the units in case no patients would be deferred to other clinics. These can be used to approximate the bed census by fitting a normal distribution with the same mean and variance. The normal approximation is included in an Integer Linear Programming (ILP) optimization model to optimize the scheduled arrivals at the clinic. Several of the assumptions are validated by means of a Discrete Event Simulation. One of the conclusions is that scheduling some patients on Saturdays smooths the bed census significantly. The authors report that their model has supported multiple clinics in the United States.

The next study focuses more on predicting the LoS of women arriving at a maternity clinic (Harper et al. 2012). The authors define a phase-type distributed LoS for two labor types: spontaneous and scheduled. For both types a decision tree based on patient characteristics, e.g. age and weight, further specifies the LoS parameters. The prediction of the LoS is then included in a simple continuous time Markov model to calculate bed occupancy for the labor ward of the clinic, using a homogeneous arrival rate. The model uses the LoS distribution and transition probabilities that women experience in each phase of labor. The steady state of the model reflects the bed census at different phases, which require different wards at the clinic.

In the literature on OBS wards we found two attempts at increasing bed occupancy, by either admitting non-OBS patients or by using 'swing rooms'. Interestingly, Harper et al. (2012) conclude that the hospital data they obtained does not show a specific time dependent arrival distribution, while others (Cochran and Bharti 2006; Griffin et al. 2012; Isken et al. 2011) do model time dependent arrival rates. Arguably, scheduled arrivals (scheduled cesarean births) likely occur only during office hours, which implies a time dependent arrival rate. Queueing models are more difficult to use in a time dependent system, since the simple formulas for waiting and blocking probability do not hold in a time dependent system. The drawback of using simulation models is that most models are case-specific, applicable only to the clinic they were designed for. However, the advantage of a graphical simulation is that practitioners can easily see the implications of different interventions, which often implies that results of the research are more easily implemented into practice. An advantage of the discrete time Markov models is that these models have the potential to mimic reality better than queueing models, and are still more general than simulation models. However, a drawback could be a rapidly increasing state space for average sized clinics consisting of multiple wards. Others propose an approximation of the bed census by a Normal distribution, and from their simulation results this seems a reasonable assumption (Isken et al. 2011).

5.3.4 Weekday Ward

Although most Dutch hospitals have a WDW and the optimization potential is significant, we were able to find only two references. This may be explained by the lack of capacity issues in these type of wards. Since all patients are elective, they can be scheduled at a time that beds are available, and patients that cannot be admitted will be accommodated on the general ward. Still, we feel that WDWs have a large logistical potential; large efficiency gains can be achieved if the number of beds is adequate and patient scheduling is optimized.

Due to the lack of modeling work on WDWs and the sparsity of scheduling work for this type of ward, we describe below two models for optimizing the patient scheduling that are relevant to the present discussion.

For a 'Monday to Friday' rheumatology clinic, admissions from a waiting list are optimized (Conforti et al. 2011). An introductory meeting determines a patient's medical priority, resource requirement and LoS. LoS is maximally 5 days. Others develop an ILP, in which they decide for each resource the patient requires (e.g. beds, diagnostic tests) at which time slot it should be scheduled, if any (Conforti et al. 2011). Each patient is assigned a weight according to his medical priority and time spend on the waiting list, while the objective is to maximize the weighted number of admissions. The authors conclude that the number of available beds is the bottleneck, and the optimized schedule can accommodate twice the number of patients compared to the schedule which was composed manually.

The last study on WDWs we found considers an online appointment scheduling version of the WDW patient scheduling problem: a patient's request arrives and should be assigned to a date and time immediately, without knowing future patient arrivals (Braaksma et al. 2015). The authors develop an Approximate Dynamic Programming model to obtain the optimal scheduling policy. This technique is often invoked when Dynamic programming models suffer from 'the curse of dimensionality', and includes aggregating the state space and approximating the value function.

5.3.5 General Ward

This section discusses models which are not applied to a specific type of ward. In most of the literature included in this section, general concepts are analyzed that are applicable to many types of wards, or the studies take multiple departments into account. Due to this generalization, most literature discussed in this section focuses on strategic or tactical planning by evaluating capacity dimensioning decisions or predicting demand.

The models for analyzing general concepts of bed census cover a wide range of OR techniques and are applied on different levels. The techniques used in the literature included in this subsection are given in Table 5.4. We will highlight

Table 5.4 Literature on general wards categorized by applied models

OR model/method	References
Algorithms	Best et al. (2015), Holm et al. (2013), van Essen et al. (2015)
Markov processes	Akkerman and Knip (2004), Gorunescu et al. (2002c), Keepers and Harrison (2009), Kusters and Groot (1996), Ramakrishnan et al. (2005), Shonick and Jackson (1973), Swain et al. (1977), Taylor et al. (2000), Utley et al. (2003a, 2005), Vasilakis et al. (2008)
Mathematical programming	Akcali et al. (2006), Bekker and Koeleman (2011), Best et al. (2015), Li et al. (2009), van Essen et al. (2015)
Queueing theory	Bekker and de Bruin (2010), Bekker and Koeleman (2011), Best et al. (2015), De Bruin et al. (2010), Gallivan and Utley (2011), Garrison and Pecina (2015), Gorunescu et al. (2002a,b,c), Green and Nguyen (2001), Griffiths et al. (2013b), Harrison et al. (2005), Li et al. (2009), Vasilakis and El-Darzi (2001), Zonderland and Boucherie (2012)
Regression	Kumar and Mo (2010)
Simulation	Akkerman and Knip (2004), Bagust et al. (1999), Dumas (1985), El-Darzi et al. (1998), Ferreira et al. (2008), Gorunescu et al. (2002c), Gunal and Pidd (2010), Harris (1986), Harrison et al. (2005), Holm et al. (2013), Keepers and Harrison (2009), Kolker (2013), Kumar (2011), Kumar and Mo (2010), Landa et al. (2014), Lapierre et al. (1999), Vanberkel and Blake (2007), Vasilakis and El-Darzi (2001), Vasilakis et al. (2008), Zhu (2011, 2014)
Stochastic models	Kortbeek et al. (2015), Mackay (2001), Vanberkel et al. (2011), Vasilakis et al. (2008)
Time series	Lapierre et al. (1999), Mackay and Lee (2005)

these models and their conclusions below by discussing a selection of the papers in Table 5.4.

A queueing model is used to determine the bed demand at community level, focusing on high occupancy rates, while keeping refusal rates of emergency patients low and waiting lists short (Shonick and Jackson 1973). The bed census is modeled using an infinite server queue incorporating two classes (elective and emergency) of arrival streams. This model elaborates on earlier research applying the infinite server queue by adding a threshold parameter (B) that blocks elective admissions if the occupancy rate is higher than or equal to B, in order to balance the elective and emergency arrival streams. This model provides policy makers useful insights in the relation between bed census, length of the waiting list and emergency refusals. Another queueing model incorporates predictable fluctuations in the average number of arrivals (Bekker and de Bruin 2010). This time-dependent queue, an $M(t)/H/s/s$ model (where $M(t)$ indicates a time-dependent Poisson process, see the Appendix for information on the notation), is evaluated by using approximations based on the infinite server queue. It is shown that daily fluctuations have limited impact on the bed census, whereas weekly patterns do have a significant

impact on both the bed census and the number of refused admissions. Finally, the authors present a method to determine the required number of beds across the week. An $M/PH/s$ queue is used to determine the optimal bed census for a hospital, in which the LoS is phase-type distributed (which is denoted by the abbreviation PH) (Gorunescu et al. 2002a). De Bruin et al. (2010) employ the Erlang loss model ($M/G/s/s$ queue) to relate the blocking probability to the occupancy. Additionally, a broad introduction of various applications of queueing networks in healthcare is also available (Zonderland and Boucherie 2012).

Several papers use a discrete Markovian approach to predict the short term bed census. These predictions are mainly based on the current bed census at day t, the expected elective and emergency admissions, and the expected discharges at day $t+j$. In these models the LoS is often empirically distributed. The census distribution is approximated from their Markov model by a Normal distribution (Utley et al. 2003a), and shows that this relatively easy approximation performs satisfactory when applied to hospital wards. Markov models are also applied to obtain the distribution of the number of patients in each phase of a care pathway, for geriatric (Taylor et al. 2000; Gorunescu et al. 2002c) or stroke patients (Vasilakis et al. 2008), in order to determine the required resources in each phase of the pathway.

Simulation is used by Dumas (1985) to analyze bed allocation and usage policies for all beds in a hospital based on hospitalizations days (e.g. 24 h bed occupancy) per specialty, average daily bed census at a certain time, bed occupancy over a time period, patient misplacements and annual misplaced patient-days. Another simulation analyzes the so called 'winter bed crisis', a yearly bed shortage during mid winter (Vasilakis and El-Darzi 2001). The results show that discharge delays during mid winter were the main reason for high bed census. The following study analyzes waiting times for surgical procedures by means of simulation (Vanberkel and Blake 2007). To balance emergency and elective admissions for the available bed capacity another simulation study was performed (Landa et al. 2014). The last simulation study focusses on the overflows between wards (in which patients are transferred to another ward because the designated ward is fully occupied), and find that the occupancy of wards is a good predictor for the frequency of overflows (Keepers and Harrison 2009).

Time series models are also used to predict bed census demand. An hourly bed sensus prediction was modelled with a time series model (Lapierre et al. 1999). The results are used to reallocate beds between different ward types such as medical, surgical or obstetric. A different but related approach involves the use of mixed exponential equations to obtain the probability distribution of patients being in different phases of their care pathways. In Mackay (2001) and Vasilakis et al. (2008) the model is applied to mimic bed census, allocating emergency admissions on both a regional and hospital level. Results show that this type of model mimics the bed occupancy accurately. The first study analyzes the accuracy of these mixed exponential equations based on a case study, and compare different equations by evaluating the effect of adding more parameters (Mackay and Lee 2005). And the latter study relates the blueprint schedule of the OT, in which each subspecialty

gets a fraction of the available OT time, to the hourly bed census distribution at the postoperative wards (Kortbeek et al. 2015).

A nonlinear mixed integer mathematical programming model is used to (re)allocate the number of available beds among different hospital services over a finite planning horizon (Akcali et al. 2006). The decisions are based on patients' waiting time before admission and budget limits. A similar technique is employed, where integer programming assists in clustering the clinical departments and assigning these clustered departments to available wards (van Essen et al. 2015). These assignments are such that capacity is sufficient to guarantee a maximum blocking probability.

Concluding, the choice for a certain modeling technique depends on the desired output. Queueing theory is suitable for determining the capacity or census distribution of a single ward, preferably with mostly unscheduled patient admissions, when a maximum blocking probability or target occupancy must be achieved. Markov models and time series models are accurate for determining the census distribution or certain percentiles, but might be tedious to analyze as the state space may become large. Simulation models can be developed as detailed or macro-leveled as desired, but are generally suitable for obtaining average performance measures. Mathematical programming can be used to optimize the reallocation of beds to wards.

5.4 Illustrations of OR Model Use

In the previous section we reviewed several OR models applied to different types of wards. In this section we provide several detailed examples of OR models applied to an ICU, OBS, AMU and WDW. All examples are based on hospital data, and illustrate the effectiveness of OR models for certain ward types. The anonymized data used for the examples is obtained from our affiliated hospitals.

5.4.1 ICU Case Study

In this case study we model the bed census of an ICU of a medium-sized Dutch teaching hospital (700 beds in total). The performance measures of interest are the bed occupancy and the probability that the bed census exceeds 40 beds (the current ICU capacity). Queueing models are therefore appropriate to apply to this case study. Hospital data shows that the number of arriving patients per day is Poisson distributed, which was expected as most patients at an ICU are urgent.

Since patients at an ICU require intensive care, deferring patients or letting them wait for a bed is not a viable solution. We therefore model the ICU with the $M/G/\infty$ queue, an infinity capacity queue, so we model the system if all patients would be accepted at the ICU. For tractability, we assume that admissions arrive according to a Poisson process.

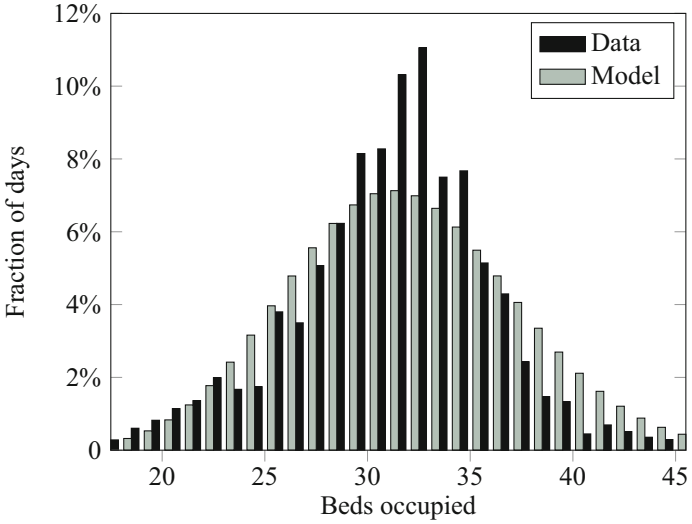

Fig. 5.2 Bed census distribution for the ICU case study

Let λ denote the arrival rate per day, $1/\mu$ the average LoS in days, and $\rho = \lambda\mu$ the load of the system. In an infinite server model the number of patients at the ward at any point in time has a Poisson distribution with parameter ρ (c.f. Winston 2003). Therefore, the probability that n beds are occupied is given by:

$$p_n = \frac{\rho^n}{n!} e^{-\rho}. \tag{5.5}$$

These probabilities are independent of the shape of the LoS distribution of patients, which is convenient for modeling an ICU as the LoS distribution at ICUs typically has a long tail, so the LoS has a high variance. In this case study the variance of the LoS is larger than the average LoS squared.

For the ICU of this case study we find from the data that the average daily arrival rate is 2.18 patients per day, and the average LoS is 14.41 days. Therefore, $\rho = 31.4$. Comparing the bed census from hospital data and the queueing model, see Fig. 5.2, we conclude that the model underestimates the probability of an 'average census' (around 32 beds) and overestimates the probability of most other census values. From the hospital data the probability that the bed census exceeds 40 equals 4.1%, while from the model this is 5.6%.

Note that in the hospital data, the bed census should not exceed the actual capacity, as this represents the realized occupancy. The fact that the census does exceed the capacity at some points in time, might be due to registration errors, for example when a nurse fills out all discharges at the end of the shift instead of the actual time of discharge or only the first ward where a patient is admitted throughout his entire stay is registered in the data (which was the case in this data). Additionally, the actual demand for beds is hard to obtain, since intensivists typically transfer a

relatively healthy patient to another ward in case of bed shortages. This complicates the validity of the model for situations close to full capacity.

With the results of this model, hospital management can gain insight about the current performance of the ICU regarding the probability that bed census exceeds capacity and occupancy. Also, the effect of management decisions can be analyzed with this model, for instance the effects of bed expansion and downscaling, or an increasing average LoS through a different patient mix.

5.4.2 OBS Case Study

In this case study we model the bed census of an OBS with 24 beds. The performance measures of interest are the bed occupancy and the probability that arriving patients have to be deferred because all beds are occupied. A queueing model is therefore an appropriate choice. Hospital data shows that the number of arriving patients per day is Poisson distributed, which was expected as most patients at an OBS are unscheduled. The hospital data shows that the arrival rate is homogeneous over the hours, except for 8:00 am; at that time most of the elective patients at the ward are admitted. For ease of modeling, we assume the arrival rate to be constant throughout the day and week. Again, we assume that admissions arrive according to a Poisson process. The performance measures of interest are best obtained by using a queue with finite capacity: an $M/G/s/s$ queue, also known as the '(Erlang) loss queue'.

Let λ denote the arrival rate per day, $1/\mu$ the average LoS in days, and $\rho = \lambda/\mu$ the average load. In the loss queue the probability that there are n patients present at a ward with capacity s beds, is given by:

$$p_n = \frac{\rho^n/n!}{\sum_{i=0}^{s} \rho^i/i!}. \tag{5.6}$$

These probabilities are independent of the LoS distribution of patients, which is in this case convenient as the LoS distribution at this OBS has a long tail.

For the OBS of this case study we find from the data that the average daily arrival rate is 9.64 patients, and the average LoS is 1.14 days. Therefore, $\rho = 10.96$. Comparing the bed census from hospital data and the queueing model, see Fig. 5.3, we conclude that the model predicts the occupancy quite accurately. The expected number of occupied beds is 10.9 according to the model, and 11.0 according to the hospital data. From the model we can determine that the probability the ward is fully occupied equals 0.025%. As the hospital does not register the number of deferred patients, we cannot verify this result.

The probability of a full ward is useful management information, since then hospital management can determine if the available capacity is still sufficient. Also this model can be used to analyze the effects on blocking probability and occupancy

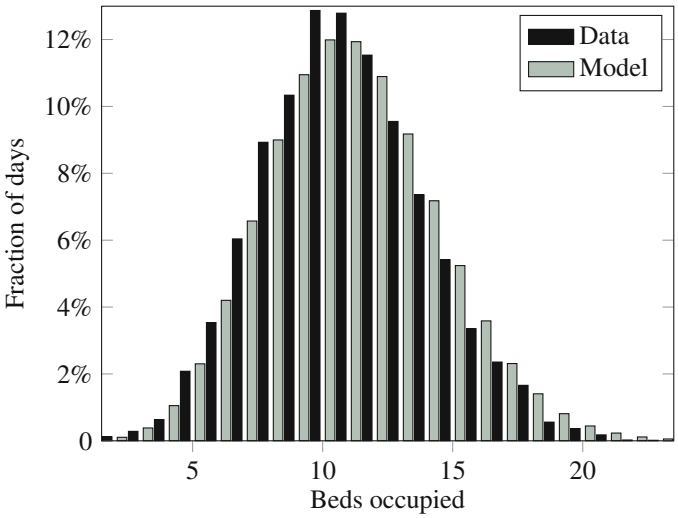

Fig. 5.3 Bed census distribution for the OBS case study

by changing the capacity of the ward. This is in fact easy to do, as there is a simple recursion between the Erlang loss probability for wards differing by one bed (i.e, server).

5.4.3 AMU Case Study

In this case study we consider a medium-sized Dutch teaching hospital (700 beds) that experiences difficulties with allocating urgent medical patients to inpatient beds, especially outside office hours. Typically, medical patients experience a longer ED LoS than surgical patients, partly due to a more complex diagnostic trajectory which involves observation of the patient and waiting until test results are available. As a result, EDs may become congested with this kind of medical patients that are under observation. Therefore, hospital management is considering opening an AMU to support the ED and medical departments. The purpose of the AMU would be faster admittance of ED patients that require observation or short hospitalization.

In preparation of the analysis, the doctors of the hospital have provided a list of diagnoses that can be admitted to the AMU. With this list the number of patients that would be admitted to the AMU if it were opened, can be estimated. Upon AMU discharge, patients either leave the hospital, or are admitted to an appropriate inpatient ward. The doctors agreed that discharges from the AMU would only occur during extended office hours (8:00 am–8:00 pm).

The performance measures of interest are the bed occupancy and blocking probability. Queueing models are therefore appropriate to apply to this case study.

As all patients are urgent and arrival rates at an ED are strongly time-dependent, we model the AMU by means of an Erlang queueing model with time-dependent arrival and service rates: an $M(t)/M(t)/s/s$ queue. Here, M_t denotes a time-dependent Poisson distribution, and s denotes the number of beds at the AMU. In a non-stationary loss queue, the limiting distribution for the number of patients in the system is time-dependent and can only be approximated. Several approximation methods exists, for example the Modified Offered Load (MOL) algorithm (Massey and Whitt 1994). Also in this case for tractability, we assume a Poisson arrival process.

The MOL algorithm approximates the load of the $M(t)/M(t)/s/s$ queue by truncating the state space of an equivalent system with infinite number of servers. Therefore the probability of having n beds occupied at time t at a ward with s beds in total, is given by:

$$P_n(t) \approx \frac{\rho(t)^n/n!}{\sum_{i=0}^{s} \rho(t)^i/i!}, \tag{5.7}$$

with $P_n(t)$ the limiting probability of n patients in the system at time t, and $\rho(t)$ the time-dependent equivalent of $\rho = \lambda/\mu$ satisfying

$$\frac{d}{dt}\rho(t) = \lambda(t) - \mu(t)\rho(t).$$

Here $\lambda(t)$ is the time-dependent arrival rate, and $\mu(t)$ is the time dependent departure rate of the AMU. The MOL approximation provides good results when the system load is moderate. In systems with high load the blocking probability is underestimated.

We obtain the time-dependent limiting probabilities of the number of occupied beds for the hospital by employing the MOL algorithm, and use these probabilities to obtain the expected bed occupancy and blocking probability. We investigate two scenarios: admitting new patients 24 h per day, or only during night time. Input for the model are the time-dependent arrival rate obtained from hospital data, depicted in Fig. 5.4, and time-dependent service rate found in hospital data. The arrival rates are adjusted to reflect the investigated scenarios. Doctors defined the patient types eligible for admitting to the AMU, and data showed that this concerned 26% of the urgent medical patients. For the hospital that commissioned this case study, opening an AMU is not warranted, as the bed occupancy would be low while the blocking probability would be high, as seen in Table 5.5. The number of patients that can be admitted to the AMU is not enough to achieve an acceptable bed occupancy and blocking probability simultaneously. Based on this results, the managers and doctors of this hospital decided not to open an AMU, and investigated other ways to reduce the ED crowding.

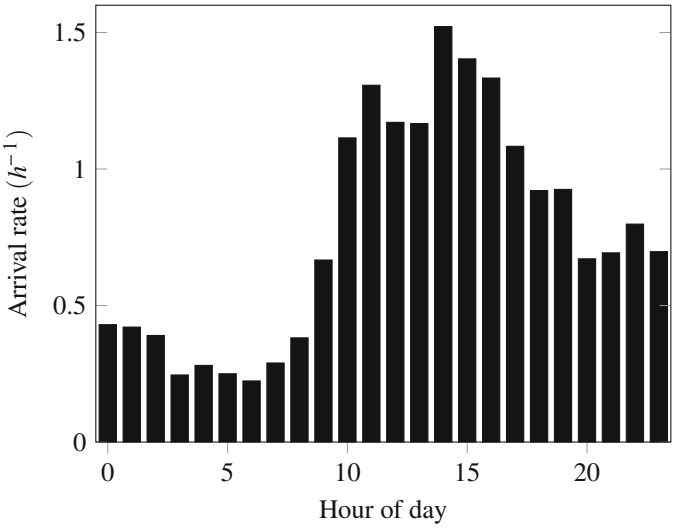

Fig. 5.4 Arrival rate per hour for urgent medical patients

Table 5.5 Results MOL algorithm, 'Occupancy' is average hourly occupancy

24 h/day admissions			Only admissions during nights		
#beds	Occupancy	Block. prob.	#beds	Occupancy	Block. prob.
5	82%	46%	3	61%	37%
8	73%	20%	4	55%	21%
10	66%	9%	6	43%	5%
12	58%	3%			

5.4.4 WDW Case Study

In this case study we optimize patient admissions to a WDW, aiming to minimize the number of beds necessary to accommodate all patients at the ward. The desired model output is both the minimum number of beds required, and a cyclic blueprint patient admission schedule. This blueprint schedule specifies for each day of the week how many of each patient type may be admitted to the ward, and there should be at least enough capacity to accommodate the average number of arriving patients.

The WDW of this case study primarily accommodates outpatients (91%), who do not stay overnight, but typically sleep off their anesthetics after a simple surgery. We assume that the cycle length is 110 h, from Monday 7:30 am until Friday 9:30 pm, as the ward closes in the weekend. Note that in this hospital, patients may be discharged after office hours. We aggregate the possible diagnoses at the WDW according to their LoS, and obtain the patient types from hospital data as given in Table 5.6.

Table 5.6 Patient types at WDW case study

Type description	LoS (h)	Av.no. patients/week
LoS < 0.5 day	5	72
0.5 day < LoS < 1 day	24	47
1 day < LoS < 1.5 day	36	8
1.5 day < LoS < 2 days	48	3

Table 5.7 Possible admission patterns WDW case study, with per LoS type the number of patients in each pattern

Pattern → LoS ↓	1	2	3	4	5	6	7	8	9	10
5	2	2	2	15		5	10	6	8	5
24		1	4		1				1	3
36		2			1	1		2	1	
48	2				1	1	1			
Sum LoS	106	106	106	75	108	109	98	102	100	97

We use an integer linear mathematical programming (ILP) model to solve this problem, which is known as the 'Cutting stock problem' (Roelofs and Bisschop 2012). We assume each bed at the WDW is available for 110 h. We define possible 'admission patterns': a combination of patients that could be placed consecutively in one bed within the opening hours of the WDW. For this case study we manually defined these patterns (see Table 5.7) as implementing too many patterns would not be practical. Note that for patterns 4, 7, 9, and 10, more patients with LoSs of 5 h could be added; we assumed the maximum number of admissions per bed per day is three, to avoid admitting patients outside office hours. The mathematical program determines the minimum number of beds necessary to accommodate all patients.

For the ILP model we need to define sets, parameters, variables, constraints, and an objective. Let $p = 1, \ldots, 11$ be the set of patterns, and $t = 1, 2, 3, 4$ the set of patient types. The parameters of the model are the demand for each patient type, D_t, and the number of patients of each type in each pattern, A_{tp}. D_t is defined in the last column of Table 5.6, and A_{tp} in Table 5.7. Define the decision variables of the ILP by x_p, the number of beds with admission pattern p. The objective is to minimize the sum of all x_p, while the constraints should reflect that all patients can be accommodated. The ILP is given by:

$$\min \sum_p x_p \tag{5.8}$$

subject to :

$$\sum_p A_{tp} x_p \geq D_t \qquad \text{for all } t \tag{5.9}$$

$$x_p \text{ integer for all } p. \tag{5.10}$$

Table 5.8 Solution of the ILP for the WDW case study

Pattern	2	3	5	9	10	11
No. beds	2	6	3	1	3	4

The ILP can be solved with many commercially available solvers, and we chose to use Microsoft Excel®. The patterns that should be used at least once and the total number of required beds are given in Table 5.8. Using this schedule the department has slots for 87 patients with LoS maximally 5 h, and precisely enough slots for the average number of the other patient types. In total 19 beds should be enough to accommodate all patients at the WDW.

The patterns are quite flexible to use in practice as the exact sequence of the patients is not specified. During admission scheduling, the WDW has to take into account that the patients should be discharged before the ward closes on Fridays.

5.5 Implemented OR Results

There exist many papers on OR models relating to different types of wards. It appears that the bed census and/or occupancy can be modeled quite accurately. However, actual use of the models in practice seems scarce; only a few of the articles reviewed for this chapter report on actual implementation results, or use of the models in practice. A widely used quote is: 'the final test of a theory is its capacity to solve the problems which originated it' (Dantzig 1963). In this section we report on the problems faced with while implementing research results, and the lessons learned from the implemented research included in this chapter.

The most important lesson from the literature is that all stakeholders (not necessarily only the problem owners) should be involved throughout the entire process to increase the likelihood of implementation (Harper and Pitt 2004; Cochran and Bharti 2006; Dumas 1985; Harper and Shahani 2002; Harris 1986; Troy and Rosenberg 2009). In the phase of defining the problem, the stakeholders determine the scope of the research, relevant performance measures, and the type of output desired, for example a new admission schedule or a decision support system. When data needs to be collected for the project, stakeholders are important for retrieving data, defining the in- and exclusion criteria, and the validation of the data. Throughout the modeling phase of the project, the stakeholders are involved in several iterations of presenting and discussing preliminary results. In the last research phase, stakeholders and/or hospital management have to decide on the recommendations they want to implement, before the actual implementation can begin.

Model input determines to a large extent the outcome and the acceptance of the results. On several occasions the already available hospital data appeared to be insufficient to provide all necessary input for the models, or the database was incomplete (Kusters and Groot 1996; Lapierre et al. 1999). Hospital data is often

inconsistent or partly missing across different databases; financial data does not always match (raw) admission and discharge data. Depending on the goals of the research, different databases may be used. Even in times of increasing use of technology, we cannot trust the data to reflect reality completely. The entry of admission and discharge data, for example, is in many hospitals still a manual task, often performed when nurses have relatively low workload or at the end of a shift. Additionally, it is important to realize that all hospital data is the *realized* process and most hospitals do not register deferred or denied patients, so actual patient demand is often hard to obtain. Knowing the ins and outs of the healthcare process is also essential in reading the data; for example, for an ICU, the LoS is affected by the bed occupancy since intensivists often transfer the healthiest patient to free a bed for a new patient in case all beds are occupied. A careful sensitivity analysis should be performed to ensure that the best possible scenario for implementation is included in the analysis.

Even when the analysis shows that one of the investigated scenarios is clearly superior to the others, a hospital may decide to implement (slightly) different interventions than those recommended. Discussions during projects often stimulate hospital staff and management to search for further possibilities for improvements (Dumas 1985; Griffin et al. 2012). During the project, a thorough robustness analysis should be performed to ensure that modified recommendations also improve the hospital's processes, and to avoid undesired outcomes and side-effects of the interventions. If possible, the interventions the hospital chooses to implement should be evaluated using the developed models.

There are two types in the implementation of research results: some hospitals implement the model, and some implement the management decision based on the results of the model. When a hospital is using the actual model, a researcher or third party develops a decision support tool that can either be included in the hospital's current software or used separately. The tool should match its user specific settings, or be flexible enough to be adapted to them (Kusters and Groot 1996; Swain et al. 1977). Additionally, users should be trained and supported in working with the model to ensure the continuation of the model's use (Harper and Shahani 2002; Swain et al. 1977).

In some projects measuring the effects of the implementations may be difficult. Hospitals may decide to implement many different interventions at the same time (Griffin et al. 2012; Kusters and Groot 1996), making it impossible to distinguish the precise impact of a particular intervention. The environment in which a ward operates may change, for example when two hospitals merge or the hospital districts are redefined (Holm et al. 2013). For prospective studies it may be unethical to measure the effect of the intervention via a randomized controlled trial, for instance: if opening an AMU appears to be the best scenario for patients, a hospital cannot set up an experiment in which one group is treated in an AMU and a control group is not. Additionally, it may also be too costly or complicated to operate a process in two different ways in parallel. Another difficulty analyzing the practical effects of an implementation is the implication of default behavior by stakeholders in models. Analyzing a system or a population, models optimize the

overall performance, while, for instance, care professionals do not act on a system level but act on individual patients. So the best option for an individual patient could be suboptimal (or even worse) for the system. Therefore the results of a modeling exercise should always be accompanied by its implications for practice. When measuring the interventions' practical effects, one should take the behavior of individuals into account.

In summary, the stakeholders play a significant role in the likelihood of implementation. Additionally, researchers should be thorough in their data collection, sensitivity and robustness analyses, and implementation support. Additional information on project life cycles for general healthcare applications is found in Harper and Pitt (2004).

5.6 Challenges and Directions for Further Research

In this chapter we have discussed various OR techniques applied to different types of wards. We elaborated on to what extent these models are implemented into practice. Some models can be applied to more than one ward type and are often used in literature. We will summarize these general models, and discuss implementation and opportunities for future research.

The most commonly applied OR techniques are queueing theory and simulation. The strength of queueing models is that straightforward formulas provide quick insights in the trade-off between occupancy and blocking probability, delay, or overflow. Simulation models can incorporate more details, but require more development time and the results are often difficult to generalize to other wards or hospitals. Using optimization models, like dynamic or mathematical programming, to analyze and optimize hospital wards seems a promising direction for future research, as literature in this direction is relatively sparse.

When it comes down to integrating the OR models into practice there is little research available. Also, the literature reviewed in this chapter does not provide much insight to what extent these models are actually implemented and/or still used in practice. This may be explained by the fact that implementation requires different competences and techniques than solely OR. We are convinced that this final and for practitioners most important phase of an OR project should receive more attention both during OR projects and in OR literature. From our own experience we know it can be challenging to make the transition from model to practice, just as doing so the other way around. Focusing on factors for successful implementation we composed the following, non-inexhaustible, list:

- Stakeholders perceive a problem
- Stakeholders are willing to and prepared for change
- The chain of command is involved
- Stakeholders are involved with every phase of the analysis
- The team defines a clear set of key performance indicators

– The team thoroughly executes data collection, model verification and validation
– The team explains practical implications of model to stakeholders
– The team takes pre and post outcome samples on the key performance indicators in order to objectively compare the effects of the implemented model in practice.

Based on the number of references per type of ward, it is also clear where the opportunities lie for OR research on wards: AMUs, OBS, and WDWs. We are confidently optimistic that this contribution guides both researchers and health care professionals through the possibilities and opportunities OR offers for wards taking trade-offs between outcomes into account.

Acknowledgements The authors would like to express their gratitude for the valuable comments of Job Kievit and Wilbert van den Hout. Additionally, we thank Aleida Braaksma for her input for one of the case studies in this project.

Appendix: OR Model Types

As background we introduce here the commonly used OR models for analyzing the performance of hospital wards. The model categories are based on the ones applied in the ORchestra database (Hulshof et al. 2011). The ORchestra database, which distinguishes the following categories: algorithms, mathematical programming, dynamic programming, regression, time series, Markov models, stochastic models, queueing theory, and simulation. We define each of the OR techniques on a basic level, and provide introductory examples.

In this section we describe the more commonly used OR models in the context of a hospital ward setting: whereas OR researchers address 'servers' we use the term 'beds', and the 'customers' are referred to as 'patients'.

Algorithms Any procedure that follows predefined steps may be called an algorithm. Algorithms are often used for solving optimization problems, and are either based upon an exact mathematical analysis, or upon some heuristic rationale. Exact algorithms return an optimal solution but have significant long runtime, while heuristics approximate the optimal solution to decrease the runtime.

Algorithms are often applied to scheduling problems. The most simple illustration of a scheduling heuristic is the 'greedy algorithm', which prescribes that we schedule every patient at the earliest available bed or appointment slot. The 'earliest due date first' heuristic schedules the patients from the waiting list at the first available resource according to ascending maximum access times. Exact algorithms are typically more complicated than heuristics, so heuristics are often preferred for practical implementations. For more information on scheduling algorithms, the reader is referred to the book of Pinedo (2015).

Mathematical Programming Mathematical programming is the name given to a variety of related fields with a common form: the optimization of one or more

objectives subject to a set of limitations, called constraints. These fields include (non-)linear (integer) programming, stochastic programming, and network flow problems. The most commonly used of these is the field of linear programming, in which the objective function and the constraints are all linear functions of the decision variables, which can be stated as follows. One seeks to optimize (that is, maximize or minimize) a single objective, which is a linear function of a vector x of decision variables (that is, variables whose values we have some control over). The solution space of x is subject to a series of linear constraints, which state the operational limitations under which the system must operate. In matrix form, a linear program to maximize the objective can be stated as:

$$\begin{aligned} &\max z = cx \\ &\text{subject to}: \quad Ax \geq b \\ &\qquad\qquad\quad x \geq 0. \end{aligned}$$

Here, c is a row vector containing the reward rates per unit increase in a particular decision variable, A is the matrix whose rows contain the coefficients for the decision variables in the various constraints, and b is the column vector of right hand sides representing the limits for these various constraints.

A more practical example of this model is given in Sect. 5.4.4. For more information, see Winston (2003). A related yet distinct area frequently used in health care applications is the field of dynamic programming, which we consider next.

Dynamic Programming All sequential decision making problems are aggregated in the dynamic programming category. This type of models break the overall decision problem into a series of more easily solved sequential problems, consisting of the different phases at which a decision maker should choose one of the available actions. In each phase the 'system' under consideration is in a certain state, where the state contains enough information to decide which action would result in the best possible outcome for the system. The chosen action may result in direct costs, and determines the state of the system in the next phase, either with certainty or known likelihood. This can be stated more formally as follows: denote the phases by t, the states by i, the possible actions by a, the direct costs associated with action a when in state i by $c(i,a)$, the probability to go from state i to j when action a is chosen by $p(i,j|a)$, and the value function $V_n(i)$. A dynamic programming model may minimize costs, or maximize rewards. A dynamic programming model (here stated with the first objective) is optimized backwards by the recursion:

$$V_n(i) = \min_a \left\{ c(i,a) + \sum_j p(j,i|a) V_{n+1}(j) \right\}.$$

Markov decision models are related to dynamic programming models. However, whereas dynamic programming works backwards in time (from phase $n+1$ to n), Markov decision problems are solved forwards in time (from phase n to $n+1$).

Dynamic programming models are therefore more suitable for problems with a given deadline, where Markov decision theory is often applied to problems with infinite horizon. For more information, see Winston (2003).

Regression and Time Series Forecasting methods are used to forecast future values of a certain variable (or variables) based on historical data. Time series models such as 'moving average' and 'exponential smoothing' take a certain number of measurements as input for the forecast. Suppose we want to estimate x_t, the average occupancy of a ward on day t. We have data on the average occupancy for each day $1, 2, \ldots, t-1$. The moving average model and exponential smoothing models are given by:

$$x_t = \frac{\sum_{i=t-N}^{t-1} x_i}{N}, \qquad A_t = \alpha x_t + (1-\alpha)A_{t-1}$$

with N the number of days used to calculate the moving average, to be determined by the user, A_t be the smoothed average of the average occupancy at day t, with $A_0 = x_0$ as starting value, and $0 < \alpha < 1$ the smoothing factor.

Regression analysis estimates the relationship between the dependent variable that we wish to forecast, x_t, and (multiple) independent variables (y_t). The linear regression model is the most simple, and is described by:

$$x_t = \beta_0 + \beta_1 y_t + \epsilon_t.$$

Here, β_0, β_1 are coefficients that set the relationship between x and y, and ϵ_t is an error term. The coefficients β_0 and β_1 should be estimated to best fit the historic occupancy, and may be determined through the least squares method.

Statistics packages such as SPSS® and Minitab® contain most forecasting tools, and also Microsoft Excel® contains formulas for forecasting. For more information, see Winston (2003) and Hamilton (1994).

Markov and Stochastic Models A stochastic model is a description of the relation between random variables, whose values are not known with certainty beforehand. A random variable measured at discrete time points, e.g. each day at 10:00, is called a discrete-time random variable. A continuous time random variable is measured continuously, for example a patient's heart rate or temperature.

Markov models are a specific type of stochastic model, and have the property that the next value in the stochastic process is independent of its past, given its current value. An example of a Markov model is the outcome of a coin toss. We use the term 'stochastic model' for all stochastic models that do not have this property and do not fall into one of the other model categories. For more information, see Ross (2007) and Winston (2003).

Queueing Theory Queueing theory is the study of waiting lines in production systems. These systems consist of a waiting line and one or multiple servers, and are defined by an arrival and service process, see Fig. 5.5.

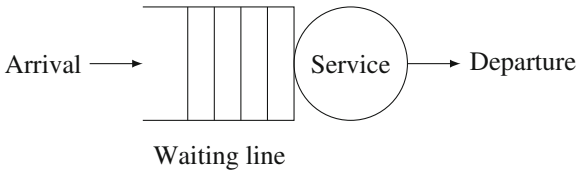

Fig. 5.5 A simple queue

A short way of referring to queues is by Kendall's notation: $A/B/s(/c)$, where A and B denote the arrival and service process, respectively, s is the number of servers, and c is an optional argument that denotes the number of places in the waiting line if this number is limited. Most queueing models assume Poisson arrivals, for which $A = M$. The service time distribution may be deterministic (D), exponential (E) or general (G). Typical performance measures that may be evaluated using queueing models are blocking probabilities, occupancy, throughput, patient waiting times, and bed idle times. Section 5.4 contains several examples of queueing models.

The QTS tool developed by Gross et al. (2008) is convenient for obtaining performance measures for most queueing (network) models with homogeneous arrival and service rates. For additional basic information on the queueing models described in this section, see Zonderland and Boucherie (2012) and Winston (2003).

Simulation Simulation models are used to mimic the evolution of a system over time, and consist of a list of what-if rules and procedures. We distinguish among discrete event simulation, Monte Carlo simulation, and system dynamics models. Discrete event simulations are event-driven routines, in which an event list is kept that contains the time stamps and types of events that will occur on those time stamps. With Monte Carlo simulation, repeated sampling from a probability distribution is carried out to obtain information on relevant performance measures. System dynamics models focus on the way different entities of the model influence each other, which relations are captured in a system of coupled, often non-linear differential equations.

Different simulation software packages exist, with different requirements regarding the user's programming abilities. Graphical simulation tools can often support the model validation process as the practitioners can see how the patients for example walk through the clinic. A drawback of graphical simulation models is that computation speed is reduced compared to non-graphical simulation packages.

For more information on simulation models, see Law (2007) and Winston (2003).

References

Akcali E, Murray JC, Lin C (2006) A network flow approach to optimizing hospital bed capacity decisions. Health Care Manag Sci 9(4):391–404

Akkerman R, Knip M (2004) Reallocation of beds to reduce waiting time for cardiac surgery. Health Care Manag Sci 7(2):119–126

Bagust A, Place M, Posnett JW (1999) Dynamics of bed use in accommodating emergency admissions: stochastic simulation model. BMJ 319(7203):155–158
Bailey NTJ (1952) A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times. J R Stat Soc Ser B (Methodological) 14(2):185–199
Bekker R, de Bruin AM (2010) Time-dependent analysis for refused admissions in clinical wards. Ann Oper Res 178(1):45–65
Bekker R, Koeleman PM (2011) Scheduling admissions and reducing variability in bed demand. Health Care Manag Sci 14(3):237–249
Best TJ, Sandikci B, Eisenstein D, Meltzer D (2015) Managing hospital bed capacity through partitioning care into focused wings. Manuf Serv Oper Manag 17(2):157–176
Blair EL, Lawrence CE (1981) A queueing network approach to health care planning with an application to burn care in New York state. Socio-Econ Plan Sci 15(5):207–216
Bountourelis T, Luangkesorn L, Schaefer A, Maillart L, Nabors SG, Clermont G (2011) Development and validation of a large scale ICU simulation model with blocking. In: Himmelspach J, White KP, Fu M, Jain S, Creasey RR (eds) Proceedings of the 2011 winter simulation conference (WSC). Intensive Care Units. IEEE, pp 1143–1153
Bountourelis T, Ulukus MY, Kharoufeh JP, Nabors SG (2012) The modeling, analysis, and management of intensive care units. International series in operations research and management science, vol 184, book section 6. Springer, New York, pp 153–182
Braaksma A, Deglise-Hawkinson J, Denton BT, Van Oyen MP, Boucherie RJ, Mes MRK (2015) Online appointment scheduling with different urgencies and appointment lengths. In: Obtained through personal communication
Broyles JR, Cochran JK, Montgomery DC (2010) A statistical Markov chain approximation of transient hospital inpatient inventory. Eur J Oper Res 207(3):1645–1657
Chan CW, Farias VF, Bambos N, Escobar GJ (2012) Optimizing intensive care unit discharge decisions with patient readmissions. Oper Res 60(6):1323–1341
Chernew ME, Newhouse JP (2012) Health care spending growth. Handb Health Econ 2:1–43
Cochran JK, Bharti A (2006) Stochastic bed balancing of an obstetrics hospital. Health Care Manag Sci 9(1):31–45
Conforti D, Guerriero F, Guido R, Cerinic MM, Conforti ML (2011) An optimal decision making model for supporting week hospital management. Health Care Manag Sci 14(1):74–88
Cooke MW, Higgins J, Kidd P (2003) Use of emergency observation and assessment wards: a systematic literature review. Emerg Med J 20(2):138–142
Costa AX, Ridley SA, Shahani AK, Harper PR, De Senna V, Nielsen MS (2003) Mathematical modelling and simulation for planning critical care capacity*. Anaesthesia 58(4):320–327
Dantzig GB (1963) Linear programming and extensions. Princeton university press, Princeton, NJ
Davies R (1994) Simulation for planning services for patients with coronary artery disease. Eur J Oper Res 72(2):323–332
de Bruin AM, van Rossum AC, Visser MCC, Koole GM (2007) Modeling the emergency cardiac in-patient flow: an application of queuing theory. Health Care Manag Sci 10(2):125–137
De Bruin AM, Bekker R, van Zanten L, Koole GM (2010) Dimensioning hospital wards using the Erlang loss model. Ann Oper Res 178(1):23–43
Dobson G, Lee H, Pinker E (2010) A model of ICU bumping. Oper Res 58(6):1564–1576
Dumas MB (1985) Hospital bed utilization: an implemented simulation approach to adjusting and maintaining appropriate levels. Health Serv Res 20(1):43–61
El-Darzi E, Vasilakis C, Chaussalet T, Millard PH (1998) A simulation modelling approach to evaluating length of stay, occupancy, emptiness and bed blocking in a hospital geriatric department. Health Care Manag Sci 1(2):143–149
Ferreira RB, Coelli FC, Pereira WC, Almeida RMVR (2008) Optimizing patient flow in a large hospital surgical centre by means of discrete-event computer simulation models. J Eval Clin Pract 14(6):1031–1037
Gallivan S, Utley M (2011) A technical note concerning emergency bed demand. Health Care Manag Sci 14(3):250–252

Garg L, McClean S, Barton M, Meenan B, Fullerton K (2010) Forecasting hospital bed requirements and cost of care using phase type survival trees. In: 2010 5th IEEE international conference intelligent systems (IS). Bed Occupancy, Cardiology, pp 185–190

Garrison GM, Pecina JL (2016) Using the M/G/∞ queueing model to predict inpatient family medicine service census and resident workload. Health Informatics Journal 22(3):429–439

Gorunescu F, McClean SI, Millard PH (2002a) A queueing model for bed-occupancy management and planning of hospitals. J Oper Res Soc 53(1):19–24

Gorunescu F, McClean SI, Millard PH (2002b) Using a queueing model to help plan bed allocation in a department of geriatric medicine. Health Care Manag Sci 5(4):307–312

Gorunescu M, Gorunescu F, Prodan A (2002c) Continuous-time Markov model for geriatric patients behavior. Optimization of the bed occupancy and computer simulation. Korean J Comput Appl Math 9(1):185–195

Green L, Nguyen V (2001) Strategies for cutting hospital beds: the impact on patient service. Health Serv Res 36:421–442

Griffin J, Xia S, Peng S, Keskinocak P (2012) Improving patient flow in an obstetric unit. Health Care Manag Sci 15(1):1–14

Griffiths JD, Knight V, Komenda I (2013a) Bed management in a critical care unit. IMA J Manag Math 24(2):137–153

Griffiths JD, Williams JE, Wood RM (2013b) Modelling activities at a neurological rehabilitation unit. Eur J Oper Res 226(2):301–312

Gross D, Shortle JF, Thompson JM, Harris CM (2008) Qtsplus software, chapter Appendix E, 4th edn. Wiley, Hoboken, NJ, pp 489–492

Gunal MM, Pidd M (2010) Discrete event simulation for performance modelling in health care: a review of the literature. J Simul 4(1):42–51

Hamilton JD (1994) Time series analysis. Princeton University Press, Princeton, NJ

Harper PR, Pitt MA (2004) On the challenges of healthcare modelling and a proposed project life cycle for successful implementation. J Oper Res Soc 55(6):657–661

Harper PR, Shahani AK (2002) Modelling for the planning and management of bed capacities in hospitals. J Oper Res Soc 53(1):11–18

Harper PR, Knight VA, Marshall AH (2012) Discrete conditional phase-type models utilising classification trees: application to modelling health service capacities. Eur J Oper Res 219(3):522–530

Harris RA (1986) Hospital bed requirements planning. Eur J Oper Res 25(1):121–126

Harrison GW, Shafer A, Macky M (2005) Modelling variability in hospital bed occupancy. Health Care Manag Sci 8(4):325–334

Holm LB, Luras H, Dahl FA (2013) Improving hospital bed utilisation through simulation and optimisation: with application to a 40 Norwegian general hospital. Int J Med Inform 82(2):80–89

Hulshof PJH, Boucherie RJ, van Essen JT, Hans EW, Hurink JL, Kortbeek N, Litvak N, Vanberkel PT, van der Veen E, Veltman B, Vliegen IMH, Zonderland ME (2011) ORchestra: an online reference database of OR/MS literature in health care. Health Care Manag Sci 14(4):383–384

Isken MW, Ward TJ, Littig SJ (2011) An open source software project for obstetrical procedure scheduling and occupancy analysis. Health Care Manag Sci 14(1):56–73

Keepers K, Harrison GW (2009) Internal flows and frequency of internal overflows in a large teaching hospital. In: McClean S, Millard P, El-Darzi E, Nugent C (eds) Intelligent patient management. Studies in computational intelligence, vol 189. Springer, Berlin/Heidelberg

Kim SC, Horowitz I, Young KK, Buckley TA (1999) Analysis of capacity management of the intensive care unit in a hospital. Eur J Oper Res 115(1):36–46

Kim SC, Horowitz I, Young KK, Buckley TA (2000) Flexible bed allocation and performance in the intensive care unit. J Oper Manag 18(4):427–443

Kokangul A (2008) A combination of deterministic and stochastic approaches to optimize bed capacity in a hospital unit. Comput Methods Prog Biomed 90(1):56–65

Kolker A (2013) Interdependency of hospital departments and hospital-wide patient flows. International series in operations research and management science, book section 2, vol 206. Springer, Boston, pp 43–63

Kortbeek N, Braaksma A, Smeenk FHF, Bakker PJM, Boucherie RJ (2015) Integral resource capacity planning for inpatient care services based on bed census predictions by hour. J Oper Res Soc 66(7):1061–1076

Kumar S (2011) Modeling hospital surgical delivery process design using system simulation: optimizing patient flow and bed capacity as an illustration. Technol Health Care 19(1):1–20

Kumar A, Mo J (2010) 'Models for bed occupancy management of a hospital in Singapore. In: Ali A (ed) Proceedings of the 2010 International conference on industrial engineering and operations management, Dhaka, pp 1–6

Kusters RJ, Groot PMA Modelling resource availability in general hospitals design and implementation of a decision support model. Eur J Oper Res 88(3):428–445

Landa P, Sonnessa M, Tanfani E, Testi A (2014) A discrete event simulation model to support bed management. In: 2014 International conference on simulation and modeling methodologies, technologies and applications (SIMULTECH). Bed Occupancy, Emergency Service, Hospital, pp 901–912

Lapierre SD, Goldsman D, Cochran R, DuBow J (1999) Bed allocation techniques based on census data. Socio-Econ Plan Sci 33(1):25–38

Law AM (2007) Simulation modeling and analysis, 4th edn. McGraw-Hill, New York

Li X, Beullens P, Jones D, Tamiz M (2009) An integrated queuing and multi-objective bed allocation model with application to a hospital in China. J Oper Res Soc 60:330–338

Litvak N, van Rijsbergen M, Boucherie RJ, van Houdenhoven M (2008) Managing the overflow of intensive care patients. Eur J Oper Res 185(3):998–1010

Mackay M (2001) Practical experience with bed occupancy management and planning systems: an Australian view. Health Care Manag Sci 4(1):47–56

Mackay M, Lee M (2005) Choice of models for the analysis and forecasting of hospital beds. Health Care Manag Sci 8(3):221–230

Mallor F, Azcarate C (2014) Combining optimization with simulation to obtain credible models for intensive care units. Ann Oper Res 221(1):255–271

Mallor F, Azcárate C, Barado J (2016) Flex Serv Manuf J 28(1–2):62–89. https://doi.org/10.1007/s10696-014-9209-8

Marcon E, Kharraja S, Smolski N, Luquet B, Viale JP (2003) Determining the number of beds in the postanesthesia care unit: a computer simulation flow approach. Anesth Analg 96(5):1415–1423

Marmor YN, Rohleder TR, Cook DJ, Huschka TR, Thompson JE (2013) Recovery bed planning in cardiovascular surgery: a simulation case study. Health Care Manag Sci 16(4):314–327

Massey WA, Whitt W (1994) An analysis of the modified offered-load approximation for the nonstationary erlang loss model. Ann Appl Probab 4(4):1145–1160

Masterson BJ, Mihara TG, Miller G, Randolph SC, Forkner ME, Crouter AL (2004) Using models and data to support optimization of the military health system: a case study in an intensive care unit. Health Care Manag Sci 7(3):217–224

McClain JO (1978) A model for regional obstetric bed planning. Health Serv Res 13(4):378–394

McManus ML, Long MC, Cooper A, Litvak E (2004) Queuing theory accurately models the need for critical care resources. Anesthesiology 100(5):1271–1276

Mustafee N, Lyons T, Rees P, Davies L, Ramsey M, Williams MD (2012) Planning of bed capacities in specialized and integrated care units: incorporating bed blockers in a simulation of surgical throughput. In: Proceedings of the 2012 winter simulation conference (WSC). IEEE, Piscataway, NJ, pp 1–12

Nguyen JM, Six P, Parisot R, Antonioli D, Nicolas F, Lombrail P (2003) A universal method for determining intensive care unit bed requirements. Intensive Care Med 29(5):849–852

Oddoye JP, Yaghoobi MA, Tamiz M, Jones DF, Schmidt P (2007) A multi-objective model to determine efficient resource levels in a medical assessment unit. J Oper Res Soc 58(12):1563–1573

Oddoye JP, Jones DF, Tamiz M, Schmidt P (2009) Combining simulation and goal programming for healthcare planning in a medical assessment unit. Eur J Oper Res 193(1):250–261
OECD. Health expenditure and financing. http://stats.oecd.org/
Pehlivan C, Augusto V, Xiaolan X, Crenn-Hebert C (2012) Multi-period capacity planning for maternity facilities in a perinatal network: a queuing and optimization approach. In: 2012 IEEE international conference on automation science and engineering (CASE). Bed Occupancy, Obstetrics, pp 137–142
Pinedo ML (2015) Scheduling: theory, algorithms, and systems, 5th edn. Springer, New York
Ramakrishnan M, Sier D, Taylor PG (2005) A two-time-scale model for hospital patient flow. IMA J Manag Math 16(3):197–215
Ridge JC, Jones SK, Nielsen MS, Shahani AK (1998) Capacity planning for intensive care units. Eur J Oper Res 105(2):346–355
Roelofs M, Bisschop J (2012) AIMMS: the user's guide, chapter 20. Paragon Decision Technology, Haarlem, pp 235–244
Ross SM (2007) Introduction to probability models, 9th edn. Academic, San Diego, CA
Scott I, Vaughan L, Bell D (2009) Effectiveness of acute medical units in hospitals: a systematic review. Int J Qual Health Care 21(6):397–407
Shahani AK, Ridley SA, Nielsen MS (2008) Modelling patient flows as an aid to decision making for critical care capacities and organisation. Anaesthesia 63(10):1074–1080
Shmueli A, Sprung CL, Kaplan EH (2003) Optimizing admissions to an intensive care unit. Health Care Manag Sci 6(3):131–136
Shonick W, Jackson JR (1973) An improved stochastic model for occupancy-related random variables in general-acute hospitals. Oper Res 21(4):952–965
Sissouras AA, Moores B (1976) The optimum number of beds in a coronary care unit. Omega 4(1):59–65
Swain RW, Kilpatrick KE, Marsh JJ (1977) Implementation of a model for census prediction and control. Health Serv Res 12(4):380–395
Taylor GJ, McClean SI, Millard PH (2000) Stochastic models of geriatric patient bed occupancy behaviour. J R Stat Soc: Ser A Stat Soc 163(1):39–48
Troy PM, Rosenberg L (2009) Using simulation to determine the need for ICU beds for surgery patients. Surgery 146(4):608–617
Utley M, Gallivan S, Treasure T, Valencia O (2003a) Analytical methods for calculating the capacity required to operate an effective booked admissions policy for elective inpatient services. Health Care Manag Sci 6(2):97–104
Utley M, Gallivan S, Davis K, Daniel P, Reeves P, Worrall J (2003b) Estimating bed requirements for an intermediate care facility. Eur J Oper Res 150(1):92–100
Utley M, Gallivan S, Jit M (2005) How to take variability into account when planning the capacity for a new hospital unit. In: Health operations management: patient flow logistics in health care, pp 146–161
van Dijk NM, Kortbeek N (2009) Erlang loss bounds for OT-ICU systems. Queueing Syst 63(1–4):253–280
van Essen JT, van Houdenhoven M, Hurink JL (2015) Clustering clinical departments for wards to achieve a prespecified blocking probability. OR Spectrum 37(1):243–271
Vanberkel P, Blake J (2007) A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. Health Care Manag Sci 10:373–385
Vanberkel PT, Boucherie RJ, Hans EW, Hurink JL, van Lent WAM, van Harten WH (2011) An exact approach for relating recovering surgical patient workload to the master surgical schedule. J Oper Res Soc 62(10):1851–1860
Vasilakis C, El-Darzi E (2001) A simulation study of the winter bed crisis. Health Care Manag Sci 4(1):31–36
Vasilakis C, El-Darzi E, Chountas P (2008) A decision support system for measuring and modelling the multi-phase nature of patient flow in hospitals. Studies in computational intelligence, book section 12, vol 109. Springer, Berlin/Heidelberg, pp 201–217

Vissers J, Beech R (2005) Health operations management: patient flow logistics in health care. Routledge, New York

Williams J, Dumont S, Parry-Jones J, Komenda I, Griffiths J, Knight V (2015) Mathematical modelling of patient flows to predict critical care capacity required following the merger of two district general hospitals into one. Anaesthesia 70(1):32–40

Winston WL (2003) Operations research: applications and algorithms, 4th edn. Brooks/Cole–Thomson Learning, Belmont, CA

Yang M, Fry MJ, Raikhelkar J, Chin C, Anyanwu A, Brand J, Scurlock C (2013) A model to create an efficient and equitable admission policy for patients arriving to the cardiothoracic ICU. Crit Care Med 41(2):414–422

Zhu Z (2011) Impact of different discharge patterns on bed occupancy rate and bed waiting time: a simulation approach. J Med Eng Technol 35(6–7):338–343

Zhu Z (2014) An online short-term bed occupancy rate prediction procedure based on discrete event simulation. J Hosp Adm 3(4):p37

Zonderland ME, Boucherie RJ, Carter MW, Stanford DA (2015) Modeling the effect of short stay units on patient admissions. Oper Res Health Care 5:21–27

Zonderland ME, Boucherie RJ (2012) Queuing networks in health care systems, book section 9. Springer, New York, pp 201–243

Part III
Care Process: Preparedness

Chapter 6
Evaluating Healthcare System Efficiency of OECD Countries: A DEA-Based Study

Zehra Önen and Serpil Sayın

6.1 Introduction

Healthcare is one of the important public policy issues of our time. Quality of healthcare services, its distribution and accessibility are often perceived as being related to quality of life of individuals. As advances in technology and sciences are reflected into healthcare domain and as demand increases, costs and expenditures have been increasing, reaching levels between 10–20% of gross domestic product (GDP) in several countries (OECD 2013). This has contributed to a higher emphasis being placed on performance evaluation for healthcare institutions (Kazandjian and Lied 1999; Özcan 2008). While healthcare organizations can be viewed and evaluated as service systems to some extent, it is difficult to assess the quality of the healthcare system of a country. In WHO (2000), the World Health Organization published a ranking of healthcare systems of 191 countries and found France as the country that has the best healthcare system. Rankings of healthcare systems of countries appear in the media from time to time as in Capell (2008), The Guardian (2003) and The New York Times (2007). Most of the time, these articles are based on WHO reports or Organisation for Economic Cooperation and Development (OECD) annual reports as presented in WHO (2015), De Looper and Lafortune (2009), and OECD (2013).

Z. Önen
Faculty of Engineering and Natural Sciences, Istanbul Bilgi University, Santralistanbul, Kazim Karabekir, Cad. No: 2/13 34060 Eyüp, İstanbul
e-mail: zonen@ku.edu.tr

S. Sayın (✉)
College of Administrative Sciences and Economics, Koç University, 34450 Sarıyer, Istanbul, Turkey
e-mail: ssayin@ku.edu.tr

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_6

The focus placed on the efficiency of the healthcare system can vary not only across countries but also across the years within a country. Different countries can have different priorities depending on the current state of their economy and other relevant parameters. In addition, when the political system changes and technology progresses, these priorities may also change.

Past research has often identified life expectancy (LE) at birth and infant mortality rate (IMR) as key outcomes of the healthcare system. Mohan and Mirmirani (2007) investigated the significance of different factors that can influence LE and IMR using panel data of 12 years, from 1990 to 2002. Two regression models were built, one where LE was chosen as the dependent variable and another where IMR was chosen as the dependent variable. Several independent variables were used in the regression models. Empirical results indicated that the level of healthcare expenditure among OECD countries has been an important factor in extending LE but did not have much impact on lowering IMR. Education level used as an indicator of health awareness was significant in both of the regressions.

Data Envelopment Analysis (DEA) has been among the methods used for performance evaluation of healthcare systems (Greene 2004; Tandon et al. 2001; Jacobs 2001). DEA is a methodology based on linear programming that can be used to analyze the relative efficiency of similar Decision Making Units (DMUs). DEA has been applied in very different areas with success (Afonso and Aubyn 2005; Galterio et al. 2009; Johnes 2006; Zhou et al. 2008). In healthcare, the methodology has been used for performance evaluation of units such as hospitals (Jacobs 2001; Özcan 2008), or different units within hospitals (Wang and Yu 2006) as well as countries (Afonso and Aubyn 2005; Reinhardt et al. 2002; Varabyova and Schreyögg 2013; Borisov et al. 2012).

Afonso and Aubyn (2005) compared performances of 24 OECD countries in education and healthcare using two different methodologies, Free Disposal Hull (FDH) and DEA. They used number of physicians, nurses and beds per 1000 population (for brevity, the term *per 1000 population* will be omitted when referring to these widely used inputs in this manuscript) as input measures and infant survival rate (ISR) and LE at birth as output measures. They used data from the year 2000. According to their results, 11 out of 24 countries were efficient in the FDH analysis whereas 8 of them remained efficient in the DEA analysis.

Greene (2004) analyzed 191 countries based on WHO data, using stochastic frontier analysis. He used disability adjusted LE and a composite measure of health care delivery as output measures. As input measures, health expenditure per capita in 1997 and average years of schooling were considered. He also used different variables that were considered as indicators of cross country heterogeneity such as the Gini coefficient that measures income inequality, and OECD membership. As a result, Greene pointed out that expenditure is a major component of healthcare system performance and should be taken into account. The author also found that OECD membership explained much of the variation in the outcome measures and the distribution of income is a significant factor.

In another study, Retzlaff-Roberts et al. (2004) developed different types of DEA models using the OECD 2000 database with data of 1998. As in Afonso

and Aubyn (2005), ISR and LE at birth were considered as output measures but these measures were dealt with separately. For input measures, they considered four healthcare-related inputs and three social environment inputs. The healthcare inputs included the number of physicians, beds, healthcare expenditure as a percentage of GDP and MRI units per million population. The social environment input variables included the expected number of years of education, the Gini coefficient and the maximum value of the percentage of male and female smokers. From their different models, the authors concluded that countries with relatively modest outcomes like Turkey and Mexico turned out to be efficient while other countries with good health outcomes were not necessarily using their resources efficiently.

Varabyova and Schreyögg (2013) analyzed the hospital care efficiency in OECD countries using panel data between 2000 and 2009. They used DEA and Stochastic Frontier Analysis (SFA) methods and compared them. For both methodologies, they considered hospital discharges and mortality as output measures whereas hospital resources such as number of beds, physicians, nurses and hospital employments were the inputs in the models. Total hospital employments refer to the number of persons employed (including self-employed and full-time equivalent employed) in general and special hospitals. Varabyova et al. argued that their analyses are good indicators of efficient use of resources. Hence, countries with good health outcomes in terms of longevity like Japan can be inefficient while developing countries like Turkey can be efficient.

Samut and Cafrı (2015) analyzed the efficiency of hospitals for 29 OECD countries between 2000 and 2010 to understand the parameters that have an impact on efficiency. They used a two stage model in which DEA and Panel Tobit were used for the first and second stage respectively. As a result they found that countries that were fully efficient during this 10-year period were Mexico, Turkey, and the United Kingdom. On the other hand, Japan, Iceland, France and Belgium had under-average efficiency scores for the same period. From the Panel Tobit analysis the authors pointed out that wealthier countries had better hospital efficiency. Indeed, they found a positive relation between GDP and education along with a positive relation between GDP and efficiency.

Frogner et al. (2015) analyzed health data of 25 OECD member countries using panel data analysis. They applied stochastic frontier analysis and fixed effect analyses over 11 input variables, including health care resources, health-related behavior, and economic and environmental factors. They built 36 different models and by comparing their results from different models, they revealed the fragility of the results of ranking models. However, they were not able to demonstrate that the U.S., in particular, performed significantly better than its WHO ranking in these alternative ranking models.

In this study, we experiment with Assurance Region Global (ARG) models based on the expectation that they may produce more conservative and consistent evaluations than those produced by standard DEA models. In addition to the widely utilized outputs of ISR and LE at birth, we also explore a different approach considering survival rates from major causes of death as output variables. Our goal is to see how efficiency and inefficiency with respect to these models differ from

those obtained by traditional models. We also provide a comprehensive evaluation of OECD member countries as some past studies exclude some member countries due to a lack of data observations (Mohan and Mirmirani 2007; Jacobs 2001; Afonso and Aubyn 2005; Retzlaff-Roberts et al. 2004; Frogner et al. 2015). We use OECD data for the years 2008 and 2012 therefore making it possible to observe any shifts in healthcare system performance at different points in time in the studied countries. In the next section, we explain the DEA methodology and present basic models. In Sect. 6.3, we describe how the data is processed, give our model results and discuss them. In Sect. 6.4, we build a model with specific causes of death as outputs. Finally, we conclude and provide further research directions in Sect. 6.5.

6.2 DEA

DEA methodology was originally introduced by Charnes, Cooper and Rhodes in 1978 as a method for evaluating the relative efficiency of Decision Making Units (DMUs) performing essentially the same task. The methodology got its name because of the idea of *enveloping* the observations to identify an efficient frontier. This frontier is computed via a ratio where multiple inputs produce multiple outputs (Joro et al. 1998; Cooper et al. 2006). The main idea of this linear programming (LP) model is to have a score between 0 and 1 representing the degree of efficiency of a DMU where 1 represents an efficient DMU. The model provides also an identification of sources and amounts of possible inefficiency and a direction of improvement based on orthogonal projection of the observation to the frontier.

Let n denote the number of DMUs to be evaluated. Suppose that there are m input and s output variables and let the input and output data be denoted by matrices X and Y of size $m \times n$ and $s \times n$ respectively. DEA methodology seeks to attach weights to each input and output variable. In the formulation below, these are decision variables denoted as $v_i,\ i = 1 \ldots, m$ and $u_j,\ j = 1, \ldots s$ respectively.

The first model introduced by Charnes, Cooper and Rhodes, denoted as CCR, sets up a fractional problem (FP_k) for an arbitrary DMU_k for $k \in \{1, \ldots, n\}$ which can be expressed as:

$$\max \theta = \frac{u_1 y_{1k} + u_2 y_{2k} + \ldots + u_s y_{sk}}{v_1 x_{1k} + v_2 x_{2k} + \ldots + v_m x_{mk}}$$

$$subject\ to \frac{u_1 y_{1j} + u_2 y_{2j} + \ldots + u_s y_{sj}}{v_1 x_{1j} + v_2 x_{2j} + \ldots + v_m x_{mj}} \leq 1, \forall j = 1, \ldots, n \tag{6.1}$$

$$v_1, \ldots, v_m \geq 0, \tag{6.2}$$

$$u_1, \ldots, u_s \geq 0, \tag{6.3}$$

The objective is to obtain the optimal weights such that the ratio of weighted output over weighted input is maximized for DMU_k. In other words, the objective is to find the most favorable weights for DMU_k that maximizes weighted output obtained per

weighted input used. Constraint (6.1) ensures that for these weights, output-to-input ratios for all DMUs are between zero and one. Constraints (6.2) and (6.3) are non-negativity constraints for the weights. Also the values for inputs and outputs are assumed to be positive. At optimality, (v_k^*, u_k^*) represents the set of most favorable weights for DMU_k that maximizes the ratio scale. Each weight shows how highly the associated input or output is evaluated relatively. Note that the above formulation is a fractional program that can be linearized relatively easily. In addition, an *output-oriented* version of this model can be written as opposed to this *input-oriented* version. The orientation is named based on the objective function of the dual problem. Discussion of equivalence between these two versions, derivations of equivalent linear programming transformations as well as a discussion of alternative DEA models can be found in Cooper et al. (2006). In this study, we employ the output oriented BCC model which differs from the CCR model by an additional convexity constraint in the dual formulation. This constraint translates into a new variable free in sign in the LP model in which the observations for the n DMUs may be combined, thus allowing a variable returns to scale in the production frontier. Formulation (BCC-O-FP) represents the dual problem with the convexity constraint and the corresponding fractional problem of an output-oriented BCC model.

$$(BCC-O-FP) \quad \min \bar{\theta} = \frac{v_1 x_{1k} + v_2 x_{2k} + \ldots + v_m x_{mk} - v_0}{u_1 y_{1k} + u_2 y_{2k} + \ldots + u_s y_{sk}}$$

$$subject\ to \frac{v_1 x_{1j} + v_2 x_{2j} + \ldots + v_m x_{mj} - v_0}{u_1 y_{1j} + u_2 y_{2j} + \ldots + u_s y_{sj}} \geq 1, \forall j = 1, \ldots, n \quad (6.4)$$

$$v_1, \ldots, v_m \geq 0, \quad (6.5)$$

$$u_1, \ldots, u_s \geq 0, \quad (6.6)$$

$$v_0 \quad free\ in\ sign \quad (6.7)$$

Note that since in BCC there is an additional constraint compared to CCR, the feasible region of the latter problem contains the feasible region of BCC. Hence, any BCC efficient DMU is CCR efficient.

DEA methodology has been used in a variety of application areas such as education, healthcare, energy efficiency, among others (Galterio et al. 2009; Johnes 2006; Zhou et al. 2008; Liu et al. 2013). One drawback of the methodology is that it may present a very favorable outlook of a DMU since it is designed to choose input and output weights in a way to benefit that particular DMU the most. In other words, the methodology is capable of overemphasizing strengths of a DMU while ignoring its weaknesses by setting respective weights to zero. As a way to alleviate this shortcoming, imposing restrictions on weight vectors u and v has been proposed. Thompson et al. (1986) developed the assurance region approach which is based on imposing constraints on the magnitude of the weights for specific inputs or outputs of DMUs relative to each other. Another proposal presented in Charnes et al. (1990), known as the cone ratio approach has been to restrict input and output weights to predetermined cones via additional constraints. Wong and Beasley (1990) propose limiting the proportion of total output of DMU k devoted to output measure i by imposing limits as follows:

$$L_i \leq \frac{u_i y_{ik}}{\sum_{j=1}^{s} u_j y_{jk}} \leq U_i. \tag{6.8}$$

Similar constraints can be defined for inputs and their corresponding weights. This approach has the advantage of restricting relative weights instead of virtual weights and therefore may be more intuitive. While Wong and Beasley (1990) motivated this approach as a means of incorporating value judgments in a DEA model, Cooper et al. (2006) mentioned that it can also be used to establish some consistency in weight choices of different DMUs by a careful choice of the bounds L_i and U_i. We employ this approach as a variation of our base model and label it as the ARG model following the terminology in Cooper et al. (2006). Since there are additional constraints in ARG models, it may be expected that some DMUs that were formerly classified as efficient may become inefficient once the constraints are imposed.

6.3 LE, Infant Mortality and Efficiency

Our data comes from the OECD online library. OECD is an organization that aims to promote policies that will improve the economic and social well-being of people around the world (OECD 2016). It has been established in 1961 and it consisted of 34 member countries at the time analyses were conducted for this study. With Latvia becoming a member in July 2016, the number of member countries has reached 35. OECD collects data from its members in order to develop policies with respect to its mission. It keeps track of significant amounts of data about not only economics, taxes, trade finance but also education, health, environment and social issues. Within healthcare domain, OECD collects a variety of data from various expenditure figures to amounts of tobacco consumption. The reader can find the list of variables related to healthcare in Health at a Glance report of OECD (2013). Table 6.1 provides a list of variables that we use in this part of our study. In these models, we pick our two output measures as ISR and LE at birth. ISR is computed using IMR as given in Afonso and Aubyn (2005) by:

$$ISR = \frac{1000 - IMR}{IMR} \tag{6.9}$$

ISR represents the ratio of children that survived their first year to the number of children that died. As inputs, we pick number of physicians, number of nurses and number of hospital beds. A vast majority of the countries account for number of physicians and nurses as practicing professionals. A few countries report the number of physicians and nurses by including practicing physicians or nurses plus others working in the healthcare sector as managers, educators and researchers, adding another 5–10% to each group. We use the figures as reported in the database. We take two cross sections of data for the 34 countries from 2008 and 2012 to be able to take two snapshots in time and to see if any differences can be observed.

Table 6.1 Description of input and output variables

Variable	Description
Number of physicians	Number of practising, professionally active or licensed to practice physicians per 1000 population.
Number of nurses	Number of practising, professionally active or licensed to practice nurses per 1000 population.
Number of hospital beds	All hospital beds regularly maintained, staffed and immediately available per 1000 population. (including curative care beds, rehabilitative care beds, long-term care beds and other hospital beds).
ISR	Infant Survival Rate. Computed via Eq. (6.9) using IMR. IMR is the number of deaths in children under 1 year of age per 1000 live births that occurred in a given year.
LE at birth	The average number of years that a person at birth is expected to live, assuming that age-specific mortality levels remain constant.

The missing data in our data set were estimated via the previous and future available data and by means of linear interpolation.[1] In the literature, it is suggested that the number of DMUs should exceed 3 times the total number of inputs and outputs. With 34 DMUs, 3 inputs and 2 outputs, we obey this guideline. All of our models are based on output-oriented BCC approach. In addition to the base models, we build ARG models where weights on outputs are imposed with the values of $L_i = 0.4$ and $U_i = 0.6$ for $i = 1, 2$ in constraint (6.8). These bounds are chosen so that nearly equal importance is given to both output variables. We use DEA-Solver-Learning Version developed by Kaoru Tone where the platform is Microsoft Excel 2003 (Cooper et al. 2006).

6.3.1 2008 Models with Respect to LE and Infant Mortality

In terms of inputs and outputs, our base BCC model is parallel to the work of Afonso and Aubyn (2005). The descriptive statistics of the variables are presented in Table 6.2. In addition to the implementation of the base model on 2008 data, we run our ARG model as well and observe the differences between the results of the two models. As we use the same inputs and outputs, in the ARG models we expect to see a subset of the efficient countries of the base model. The efficiency scores of all countries for both BCC and ARG models can be found in Table 6.3. In the following tables with efficiency scores, a score of 1 is presented in bold in order to highlight countries that are efficient. In 2008 BCC model, Canada, Chile, Greece, Italy, Japan, South Korea, Luxembourg, Mexico, New Zealand, Spain, Sweden and Switzerland

[1] A total of 14 data items, 8 for 2008, 6 for 2012 of the total 340 have been estimated.

Table 6.2 Descriptive statistics of variables for 2008

	Number of physicians	Number of nurses	Number of beds	LE	ISR
Mean (σ)	3.04 (0.91)	8.35 (3.91)	5.08 (2.39)	79.24 (2.55)	269.8 (104.2)
Max (DMU)	6 (Greece)	14.9 (Switzerland)	13.8 (Japan)	82.7 (Japan)	554.6 (Luxembourg)
Min (DMU)	1 (Chile)	0.5 (Chile)	1.7 (Mexico)	73.6 (Turkey)	64.8 (Mexico)

Table 6.3 Scores of all OECD countries for 2008 models

DMU	BCC	ARG	DMU	BCC	ARG
Australia	0.996	0.776	Japan	**1**	0.992
Austria	0.980	0.769	South Korea	**1**	0.994
Belgium	0.972	0.699	Luxembourg	**1**	**1**
Canada	**1**	0.801	Mexico	**1**	**1**
Chile	**1**	**1**	Netherlands	0.981	0.772
Czech Republic	0.945	0.872	New Zealand	**1**	0.973
Denmark	0.965	0.756	Norway	0.992	0.929
Estonia	0.903	0.682	Poland	0.936	0.727
Finland	0.980	0.886	Portugal	0.974	0.955
France	0.987	0.771	Slovak Republic	0.916	0.627
Germany	0.975	0.727	Slovenia	0.983	0.975
Greece	**1**	**1**	Spain	**1**	**1**
Hungary	0.903	0.652	Sweden	**1**	**1**
Iceland	0.997	0.869	Switzerland	**1**	0.693
Ireland	0.978	0.728	Turkey	0.940	0.632
Israel	0.995	0.910	United Kingdom	0.982	0.786
Italy	**1**	0.934	United States	0.964	0.698

are efficient. We can see that some of the developed economies (e.g. Germany) are inefficient whereas some of the developing economies are efficient (e.g. Chile).[2] A similar counter intuitive result was pointed out in the work of Retzlaff-Roberts et al. (2004) with data from OECD 2000 database. Note that due to the nature of DEA methodology, inefficiency does not necessarily imply a deficiency in outputs. The developed countries may be using more inputs compared to developing ones for getting certain level of outputs whereas developing countries do relatively well with their limited available resources. A closer look at the output weights of the efficient

[2]The distinction between developing and developed economies is made based on the classification of the UN (2015) World Economic Situation and Prospect report.

Table 6.4 Descriptive statistics of variables for 2012

	Number of Physicians	Number of Nurses	Number of Beds	LE	ISR
Mean (σ)	3.22 (0.93)	8.97 (4.07)	4.80 (2.50)	80.18 (2.44)	307.2 (151.1)
Max (DMU)	6.27 (Greece)	16.97 (Switzerland)	13.36 (Japan)	83.2 (Japan)	908.1 (Iceland)
Min (DMU)	1.73 (Turkey)	1.79 (Turkey)	1.57 (Mexico)	74.4 (Mexico)	74.2 (Mexico)

countries reveals that in 2008, Greece, Luxembourg, Spain and Sweden have zero weight on their LE at birth output. 16 out of the 22 inefficient countries achieve their best by nullifying one of their output weights.

In the 2008 ARG model, six of the BCC-efficient countries (Chile, Greece, Luxembourg, Mexico, Spain and Sweden) are still efficient. On the other hand, Canada, Italy, Japan, South Korea, New Zealand and Switzerland lost their efficiency once the additional constraints are added. We note that countries that lose their efficiency when a more balanced combination of outputs is enforced are developed countries. This strengthens the interpretation that the inefficiencies may stem from abundant inputs rather than poor outputs. We also observe that scores now come from a wider range between 0.627 and 1, as expected.

6.3.2 2012 Models with Respect to LE and Infant Mortality

The descriptive statistics of the variables for 2012 are presented in Table 6.4. In the 2012 BCC model, some countries that were efficient in 2008 are not efficient whereas some countries that were not efficient in 2008 are observed as efficient in 2012. The countries that were efficient in 2008 but not in 2012 are Italy, Luxembourg, New Zealand and Switzerland. In 2012, Iceland, Israel, Slovenia and Turkey join the list of efficient countries. Even more inefficient countries (21 out of 22) have zero weight on one of their output weights. In 2012, Canada, Israel and Sweden are the countries that are BCC efficient but inefficient with respect to the ARG model. Again, it is mainly developed countries that are forced to inefficiency by a balanced use of outputs. Furthermore in the ARG model scores vary between 0.523 and 1. All the results for BCC and ARG models for 2012 can be found in Table 6.5.

Table 6.5 Scores of all OECD countries for 2012 models

DMU	BCC	ARG	DMU	BCC	ARG
Australia	0.995	0.673	Japan	**1**	**1**
Austria	0.978	0.730	South Korea	**1**	**1**
Belgium	0.976	0.611	Luxembourg	0.990	0.777
Canada	**1**	0.676	Mexico	**1**	**1**
Chile	**1**	**1**	Netherlands	0.981	0.565
Czech Republic	0.945	0.794	New Zealand	0.993	0.575
Denmark	0.969	0.572	Norway	0.982	0.657
Estonia	0.928	0.756	Poland	0.950	0.730
Finland	0.976	0.751	Portugal	0.975	0.864
France	0.991	0.658	Slovak Republic	0.925	0.591
Germany	0.975	0.598	Slovenia	**1**	**1**
Greece	**1**	**1**	Spain	**1**	**1**
Hungary	0.913	0.623	Sweden	**1**	0.832
Iceland	**1**	**1**	Switzerland	0.997	0.523
Ireland	0.991	0.732	Turkey	**1**	**1**
Israel	**1**	0.970	United Kingdom	0.992	0.722
Italy	0.997	0.915	United States	0.966	0.565

6.3.3 Discussion of Results

Looking at the overall results, it is possible to observe that Chile, Greece, Mexico and Spain are efficient in both models for both years. On the other hand, 18 countries (including Austria, France, the United Kingdom and the United States) out of 34 are always inefficient regardless of the model types and years.

We conducted reference set analyses on the 2008 and the 2012 ARG models. Tables 6.6 and 6.7 report the reference set members along with the associated weights[3] for the inefficient countries for 2008 and 2012, respectively. In Table 6.6 we can see that Luxembourg is dominant with an occurrence of 26 out of 28. Furthermore, Luxembourg is the country with the highest reference set weight value for more than half of the countries. Therefore it can be stated that Luxembourg, with the highest ISR (554.6) and its LE at birth being relatively good (80.7), can be seen as the "ideal" country in terms of healthcare system performance. This outcome is worth attention because Retzlaff-Roberts et al. did not include Luxembourg in their data set due to lack of information in OECD 2000 database. Chile, Sweden and Spain are the other countries that appear frequently in the reference sets. An interesting observation is that Luxembourg is not efficient in either of the 2012 models although it has a strong appearance in 2008 models. DEA model results include the values of the slacks that give a direction of improvement from the computation of the model. We remark that most of the excess is present in the number of nurses and number of hospital beds according to ARG models.

[3]Weight are rounded to two digit accuracy and the total may deviate from 1.00 for some countries.

Table 6.6 Reference set for inefficient countries for ARG model in 2008

DMU	Reference set
Australia	Sweden (0.49); Luxembourg (0.38); Chile (0.13)
Austria	Luxembourg (0.56); Greece (0.44)
Belgium	Luxembourg (1.00)
Canada	Chile (0.43); Sweden (0.31); Luxembourg (0.26)
Czech Republic	Luxembourg (0.64); Greece (0.27); Chile (0.09)
Denmark	Sweden (0.70); Luxembourg (0.29); Chile (0.01)
Estonia	Luxembourg (0.49); Greece (0.28); Chile (0.22)
Finland	Luxembourg (0.87); Chile (0.1); Greece (0.03)
France	Luxembourg (0.65); Greece (0.22); Chile (0.12)
Germany	Luxembourg (0.97); Greece (0.03)
Hungary	Luxembourg (0.48); Chile (0.27); Greece (0.25)
Iceland	Luxembourg (1.00)
Ireland	Luxembourg (0.76); Sweden (0.2); Chile (0.04)
Israel	Spain (0.68); Luxembourg (0.14); Chile (0.09); Greece (0.09)
Italy	Spain (0.53); Luxembourg (0.3); Chile (0.12); Sweden (0.04)
Japan	Luxembourg (0.67); Chile (0.33)
South Korea	Chile (0.59); Luxembourg (0.36); Greece (0.05)
Netherlands	Luxembourg (0.65); Spain (0.29); Chile (0.06)
New Zealand	Sweden (0.42); Mexico (0.35); Chile (0.22)
Norway	Sweden (0.75); Luxembourg (0.25)
Poland	Chile (0.48); Luxembourg (0.43); Greece (0.09)
Portugal	Spain (0.89); Luxembourg (0.09); Sweden (0.02)
Slovak Republic	Luxembourg (0.47); Greece (0.31); Chile (0.22)
Slovenia	Luxembourg (0.7); Chile (0.27); Greece (0.03)
Switzerland	Luxembourg (0.86); Sweden (0.14)
Turkey	Chile (0.86); Spain (0.08); Sweden (0.05)
United Kingdom	Sweden (0.41); Chile (0.32); Luxembourg (0.27)
United States	Chile (0.42); Sweden (0.4); Luxembourg (0.18)

A reference set analysis for 2012 is given in Table 6.7. It can be observed that Slovenia is dominant with an occurrence of 21 out of 25 times. Also Iceland has an occurrence of 17 out of 25. Chile, Greece, Spain and Turkey are the other countries that appear frequently in the reference sets.

6.4 Survival from Major Causes of Death and Efficiency

Our results above show that when LE at birth and ISR are taken as the outputs that indicate system efficiency, developing countries may have an advantage as they achieve relatively good results in relation to the inputs they provide to the system.

Table 6.7 Reference set for inefficient countries for ARG model in 2012

DMU	Reference set
Australia	Slovenia (0.42); Iceland (0.38); Spain (0.2)
Austria	Slovenia (0.93); Greece (0.07)
Belgium	Slovenia (0.81); Iceland (0.19)
Canada	Chile (0.57); Iceland (0.4); Slovenia (0.03)
Czech Republic	Slovenia (0.98); Greece (0.02)
Denmark	Iceland (0.91); Mexico (0.09)
Estonia	Slovenia (0.54); Greece (0.23); South Korea (0.18); Turkey (0.05)
Finland	Slovenia (0.54); Iceland (0.46)
France	Slovenia (0.87); Iceland (0.13)
Germany	Iceland (0.64); Slovenia (0.36)
Hungary	Slovenia (0.51); South Korea (0.3); Greece (0.18)
Ireland	Iceland (0.53); Chile (0.46); Slovenia (0.01)
Israel	Spain (0.67); Turkey (0.21); Slovenia (0.11); Greece (0.01)
Italy	Spain (0.69); Slovenia (0.27); Iceland (0.04)
Luxembourg	Slovenia (0.77); Iceland (0.23)
Netherlands	Iceland (0.57); Slovenia (0.43)
New Zealand	Iceland (0.51); Chile (0.46); Slovenia (0.03)
Norway	Iceland (1.00)
Poland	South Korea (0.42); Slovenia (0.39); Turkey (0.18); Greece (0.01)
Portugal	Spain (0.73); Slovenia (0.26); Greece (0.01)
Slovak Republic	Slovenia (0.44); Greece (0.26); South Korea (0.24); Turkey (0.05)
Sweden	Iceland (0.62); Mexico (0.38)
Switzerland	Iceland (1.00)
United Kingdom	Mexico (0.42); Iceland (0.35); Slovenia (0.21); Spain (0.02)
United States	Chile (0.49); Iceland (0.35); Slovenia (0.16)

Obviously our models correspond to a very high level analysis of the healthcare system. In order to focus on the system's ability to deal with major health issues, we analyze some major causes of death and their corresponding survival rates as outputs. According to WHO fact sheet, the first two major causes of death in the world are ischemic heart disease and stroke. Death from all types of cancer is also listed as a major cause of death in WHO (2014). We build new models by using survival rates of ischemic heart disease, cerebrovascular disease (since stroke is the most common type of cerebrovascular diseases) and malignant neoplasms as outputs of the system while the inputs remain the same. We label our new models that consider survival rates from three major causes of death (ischemic heart disease, cerebrovascular disease, malignant neoplasms) as outputs along with the same inputs as in previous models BCCs and ARGs. Data from OECD library is presented in the form of age standardized mortality rates per 100,000 population. These rates are calculated by the OECD Secretariat, using the total OECD population for 2010 of each corresponding country as the reference population. They use the method

Table 6.8 Descriptive statistics of mortality rates from common causes of death for 2008

	Ischemic heart MR	Cerebrovascular MR	Malignant neoplasms MR
Mean (σ)	135.72 (86.31)	78.75 (29.35)	215.06 (32.66)
Max (Country)	428.4 (Slovak Republic)	148.4 (Slovak Republic)	294.7 (Hungary)
Min (Country)	41.9 (Japan)	41.8 (Israel)	124.5 (Mexico)

of standardization for age-standardized calculations to be able to compare the level of mortality across countries and over time. We compute survival rates based on mortality rates using Eq. (6.9) replacing 1000 by 100,000.

6.4.1 2008 Models with Respect to Survival from Major Causes of Death

For 2008, Turkey is the only country with missing data. We use data from 2009 for Turkey. The descriptive statistics of the output variables for 2008 are presented in Table 6.8.

Table 6.9 shows the results of the BCC and ARG models. In the BCC model nearly half of the countries (16 out of 34) are efficient. The efficient countries are Canada, Chile France, Israel, Italy, Japan, South Korea, Luxembourg, Mexico, Netherlands, Portugal, Slovenia, Switzerland, Turkey, the United Kingdom and the United States. Compared to the BCC model in Sect. 6.3.1, France, Israel, Netherlands, Portugal, Slovenia, Turkey, the United Kingdom and the United States became efficient with the new output variables. On the other hand, Greece, New Zealand and Sweden are no longer efficient. Looking at the ARG model results, we see that Portugal, Slovenia, Turkey and the United Kingdom are not listed as efficient anymore. Also looking at the scores, we see that the discrepancy is much more apparent when the output variables changed. This can be explained by high standard deviation values of the output variables.

6.4.2 2012 Models with Respect to Survival From Major Causes of Death

The most recent and complete data about survivals from the three conditions belong to 2012. Data on survivals from these conditions are not available for Canada, Iceland and Slovenia. We use the most recent data for these countries: 2011 for Canada, 2010 for Iceland and 2009 for Slovenia. The descriptive statistics of the variables are presented in Table 6.10. Table 6.11 shows the results of both models for BCC and ARG models using 2012 data.

Table 6.9 Scores of all OECD countries for 2008 mortality output models

	BCCs	ARGs		BCCs	ARGs
Australia	0.794	0.777	Japan	**1**	**1**
Austria	0.812	0.673	South Korea	**1**	**1**
Belgium	0.794	0.791	Luxembourg	**1**	**1**
Canada	**1**	**1**	Mexico	**1**	**1**
Chile	**1**	**1**	Netherlands	**1**	**1**
Czech Republic	0.505	0.417	New Zealand	0.783	0.745
Denmark	0.773	0.773	Norway	0.781	0.781
Estonia	0.547	0.412	Poland	0.558	0.554
Finland	0.784	0.631	Portugal	**1**	0.845
France	**1**	**1**	Slovak Republic	0.494	0.373
Germany	0.778	0.700	Slovenia	**1**	0.642
Greece	0.737	0.701	Spain	**1**	**1**
Hungary	0.450	0.334	Sweden	0.766	0.766
Iceland	0.804	0.691	Switzerland	**1**	**1**
Ireland	0.694	0.619	Turkey	**1**	0.903
Israel	**1**	**1**	United Kingdom	**1**	0.688
Italy	**1**	**1**	United States	**1**	**1**

Table 6.10 Descriptive statistics of mortality rates from common causes of death for 2012

	Ischemic heart MR	Cerebrovascular MR	Malignant neoplasms MR
Mean (σ)	114.32 (71.79)	66.27 (23.92)	207.19 (32.03)
Max (Country)	318.6 (Slovak Republic)	122.9 (Slovak Republic)	293.3 (Hungary)
Min (Country)	37.7 (Japan)	37.1 (Switzerland)	117.1 (Mexico)

We observe that Canada, Chile, France, Israel, Japan, South Korea, Mexico, Spain and Turkey are efficient in both models. Portugal and Switzerland are BCC efficient but ARG inefficient. France, Portugal and Switzerland are efficient when survival rates from major causes of death are considered as outputs instead of LE at birth and ISR. On the other hand, Greece, Iceland Slovenia and Sweden are not efficient when survival rates from major causes of death are considered as outputs although they are efficient with respect to LE at birth and ISR. Again, we observe that efficiency and inefficiency do not necessarily align with the status of countries as developed or developing. Looking at the reference sets presented in Table 6.12, even though France is not efficient when classical outputs are considered, this country can be followed as the model country for most of the inefficient countries of the ARG model when main causes of mortality are considered as output variables. Also Canada, Mexico and Spain appear frequently in the reference sets.

Table 6.11 Scores of all OECD countries for 2012 mortality output models

	BCCs	ARGs		BCCs	ARGs
Australia	0.868	0.856	Japan	**1**	**1**
Austria	0.833	0.653	South Korea	**1**	**1**
Belgium	0.829	0.826	Luxembourg	0.865	0.865
Canada	**1**	**1**	Mexico	**1**	**1**
Chile	**1**	**1**	Netherlands	0.921	0.902
Czech Republic	0.526	0.414	New Zealand	0.717	0.713
Denmark	0.865	0.849	Norway	0.777	0.773
Estonia	0.659	0.476	Poland	0.598	0.597
Finland	0.742	0.622	Portugal	**1**	0.853
France	**1**	**1**	Slovak Republic	0.465	0.381
Germany	0.790	0.699	Slovenia	0.628	0.611
Greece	0.802	0.768	Spain	**1**	**1**
Hungary	0.431	0.359	Sweden	0.804	0.804
Iceland	0.724	0.702	Switzerland	**1**	0.944
Ireland	0.761	0.752	Turkey	**1**	**1**
Israel	**1**	**1**	United Kingdom	0.794	0.794
Italy	0.767	0.760	United States	0.961	0.910

6.5 Conclusion

This research aims to provide an evaluation of healthcare system efficiency of 34 OECD countries. DEA methodology is used with different modeling techniques and with different output measures. OECD data for 2008 and 2012 are used. The two output measures we use in the first models, LE at birth and ISR, are generally accepted measures as system output in the literature. We suggest the use of ARG models in a way to balance the two outputs as a remedy against the overly optimistic nature of the DEA methodology. To our knowledge, this is the first study that suggests using an ARG model for evaluating healthcare system efficiency at the country level. We observe a more consistent result when constraints on relative weights are imposed as weight restrictions require a more balanced relative output generation. A reference set analysis on the ARG model of 2008 displays Luxembourg as a good role model for inefficient countries whereas for 2012, Slovenia assumes that role.

In addition to using LE and ISR as outputs as is traditionally done, we experiment with survival rates from major causes of death as possible measures of outputs and implement this model with 2008 and 2012 data. When compared to LE at birth as a measure, this model rules out deaths from less common causes as well as deaths whose occurrence have a more indirect relationship to the performance of the healthcare system, such as deaths from accidents, homicides and suicides. In other words, under the assumption that most countries spend more effort on increasing their healthcare system capacity geared towards conditions surrounding

Table 6.12 Reference set for inefficient countries for ARGs model in 2012

DMU	Reference set
Australia	France (0.38); Canada (0.35); Mexico (0.28)
Austria	France (0.67); Mexico (0.22); Japan (0.11)
Belgium	France (0.65); Mexico (0.19); Japan (0.14); Canada (0.02)
Czech Republic	France (0.79); South Korea (0.15); Mexico (0.06)
Denmark	Spain (0.82); Canada (0.1); France (0.05); Mexico (0.02)
Estonia	France (0.47); South Korea (0.19); Mexico (0.29); Israel (0.05)
Finland	France (0.74); Mexico (0.24); Japan (0.02)
Germany	France (0.72); Japan (0.28)
Greece	Turkey (0.3); South Korea (0.3); Spain (0.22); Mexico (0.18)
Hungary	France (0.44); South Korea (0.38); Mexico (0.18)
Iceland	Mexico (0.42); Canada (0.29); France (0.29)
Ireland	Canada (0.65); Mexico (0.25); France (0.1)
Italy	Spain (0.37); Mexico (0.35); France (0.28)
Luxembourg	France (0.58); Mexico (0.18); Chile (0.16); Japan (0.06); Canada (0.03)
Netherlands	France (0.58); Spain (0.26); Mexico (0.16)
New Zealand	Mexico (0.47); Canada (0.35); France (0.18)
Norway	France (0.46); Mexico (0.37); Canada (0.17)
Poland	Mexico (0.61); Japan (0.31); Canada (0.05); France (0.03)
Portugal	Spain (0.87); France (0.13)
Slovak Republic	France (0.4); Mexico (0.32); South Korea (0.28)
Slovenia	France (0.4); Mexico (0.33); Chile (0.19); Japan (0.08)
Sweden	Mexico (0.44); Canada (0.26); Israel (0.2); France (0.09)
Switzerland	France (0.61); Canada (0.28); Mexico (0.11)
United Kingdom	Canada (0.48); Spain (0.22); Mexico (0.18); France (0.08); Chile (0.05)
United States	Canada (0.71); Mexico (0.12); France (0.11); Chile (0.06)

most likely causes of death, efficiency with respect to this model can be taken as an indication of their success. We observe that while there is significant overlap among efficient countries of different models, there are also differences. Countries that are not efficient with respect to LE and ISR but are efficient with respect to survival rates from major causes of death may focus on other causes of death to improve their standing. On the other hand, countries that are efficient with respect to LE and ISR but are not efficient with respect to survival rates from major causes of death need to focus more on these major causes of death. A reference set analysis on the ARG model of 2008 and 2012 displays France as a good role model for inefficient countries with these two outputs whereas for the previous models.

Lack of output measures that capture the quality of health services remains as a limitation of the current study. This is mainly due to the difficulty of quantifying system quality and obtaining associated complete data. Bringing commonly accepted appropriate pseudo-measures that quantify system quality into our models remains as future work. Another future research direction is to study how efficiency

of healthcare systems change over time by using, for instance, the Malmquist Index, named after Malmquist (1953). This may be more revealing in terms of the nature of the changes that take place in the studied time interval. Although we studied two snapshots in time to observe any possible shifts with respect to different DEA models and different output measures and also to better relate to past literature, a multi-year study designed to track changes in system performance might provide more insights to policy makers.

References

Afonso A, Aubyn MS (2005) Non-parametric approaches to education and health efficiency in OECD countries. J Appl Econ 8(2):227–246

Borisov D, Cicea C, Turlea C (2012) DEA model for assessing efficiency in providing health care and management decisions. Manage Res Pract 4(1):5–18

Capell K (2008) The french lesson in health care. The Guardian, 26 June 2008. http://www.guardian.co.uk/environment/2008/jun/26/climatechange.scienceofclimatechange

Charnes A, Cooper WW, Huang ZM (1990) Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks. J Econometrics 46(1–2):73–91

Cooper WW, Seiford LM, Tone K (2006) Introduction to data envelopment analysis and its uses with DEA-solver software and references, Springer, Berlin

De Looper M, Lafortune G (2009) Measuring disparities in health status and in access and use of health care in OECD countries. Organisation for Economics Co-operation and Development

Frogner BK, Frech H, Parente ST (2015) Comparing efficiency of health systems across industrialized countries: a panel analysis. BMC Health Serv Res 15(1):1

Galterio L, Helton J, Langabeer J, 2nd, DelliFraine J (2009) Data envelopment analysis: performance normalization and benchmarking in healthcare. J Health Inf Manag 23(3):38–43

Greene W (2004) Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the world health organization's panel data on national health care systems. Health Econ 13(10):959–980

Jacobs R (2001) Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis. Health Care Manage Sci 4:103–115

Johnes J (2006) Data envelopment analysis and its application to the measurement of efficiency in higher education. Econ Educ Rev 25(3):273–288

Joro T, Korhonen P, Wallenius J (1998) Structural comparison of data envelopment analysis and multiple objective linear programming. Manage Sci 44(7):962–970

Kazandjian VA, Lied T (1999) Healthcare performance measurement: systems design and evaluation. ASQ Quality Press, Wisconsin

Liu JS, Lu LY, Lu WM, Lin BJ (2013) A survey of dea applications. Omega 41(5):893–902

Malmquist S (1953) Index numbers and indifference surfaces. Trabajos de Estadistica y de Investigacion Operativa 4(2):209–242

Mohan R, Mirmirani S (2007) An assessment of OECD health care system using panel data analysis. Southwest Bus Econ J 16:21–35

OECD (2013) Health at a glance 2013: OECD indicators. OECD

OECD (2016) Secretary-general's report to ministers 2016. http://dx.doi.org/10.1787/sg_report-2016-en

Özcan YA (2008) Health care benchmarking and performance evaluation: an assessment using data envelopment analysis. International Series in Operations Research and Management Science, vol 120. Springer, Berlin

Reinhardt UE, Hussey PS, Anderson GF (2002) Cross-national comparisons of health systems using OECD data, 1999. Health Affair 21(3):169–181

Retzlaff-Roberts D, Chang FC, Rubin RM (2004) Technical efficiency in the use of health care resources: a comparison of OECD countries. Health Policy 69:55–72

Samut PK, Cafrı R (2015) Analysis of the efficiency determinants of health systems in OECD countries by DEA and panel Tobit. In: Social indicators research. Springer, Berlin, pp 1–20

Tandon A, Murray CJ, Lauer JA, Evans DB (2001) Measuring overall health system performance for 191 countries. World Health Organization

The Guardian (2003) NHS slips down world healthcare league. The Guardian-Society, 14 November 2003. http://www.theguardian.com/society/2003/nov/14/politics.medicineandhealth

The New York Times (2007) World's best medical care? The New York Times, 12 August 2007. http://www.nytimes.com/2007/08/12/opinion/12sun1.html?pagewanted=all&_r=0

Thompson RG, Singleton FDJ, Thrall RM, Smith BA (1986) Comparative site evaluations for locating a high-energy physics lab in Texas. Interfaces 16:35–49

UN (2015) World economic situation and prospects statistical annex 2015. http://www.un.org/en/development/desa/policy/wesp/wesp_archive/2015wespstats-en.pdf

Varabyova Y, Schreyögg J (2013) International comparisons of the technical efficiency of the hospital sector: panel data analysis of OECD countries using parametric and non-parametric approaches. Health Policy 112(1):70–79

Wang R, Yu M (2006) Evaluating the efficiency of hospital's departments using DEA. In: Proceedings of IEEE international conference on service operations and logistics, and informatics (SOLI 2006), pp 1167–1170

WHO (2000) World health organization (2000) the world health report 2000: health systems: improving performance. http://www.who.int/whr/2000/en/

WHO (2014) The top 10 causes of death. Fact sheet no 310, May 2014. http://www.who.int/mediacentre/factsheets/fs310/en/

WHO (2015) World Health Organization (2015) world health statistics 2015. http://www.who.int/gho/publications/world_health_statistics/2015/en/

Wong YHB, Beasley JE (1990) Restricting weight flexibility in data envelopment analysis. J Oper Res Soc 41(9):829–835

Zhou P, Ang BW, Poh KL (2008) A survey of data envelopment analysis in energy and environmental studies. Eur J Oper Res 189(1):1–18

Chapter 7
Healthcare Expenditure Prediction in Turkey by Using Genetic Algorithm Based Grey Forecasting Models

Tuncay Özcan and Fatih Tüysüz

7.1 Introduction

Forecasting can be defined as the process of making predictions about future mainly based on past and present data and analysis of trends. The reason for forecasting is to cope with the uncertainty of future by providing information about probable events and their consequences. Forecasting is essential in decision making and can guide decisions in a variety of fields such as operation planning and control, marketing, economics, finance, capacity planning, business and government budgeting, demography, crisis management and etc. (Diebold 1998).

There are many methods and techniques developed for forecasting. There is also different classification of the techniques in literature. Mun (2006) classifies forecasting techniques generally as quantitative and qualitative approaches. Qualitative forecasting is used when no reliable or relevant past data exists. Qualitative methods are the Delphi or expert opinion approach, management assumptions, market research or external data and surveys. Quantitative methods divide the available data into time-series (ARIMA, multivariate regression, nonlinear regression, stochastic processes), cross-sectional (econometric model, Monte Carlo simulation, multiple regression), or mixed panel (ARIMA (X), multiple regression). Chatfield (2000) broadly classifies forecasting methods into three types which are judgemental forecasts based on subjective judgement and intuition, univariate methods where forecasts are based on present and past values of the single series being forecasted, and multivariate methods where forecasts for a variable are based on values of one or more time series variables. In another point of view, Xia and Wong (2014) divide the forecasting techniques into two groups which are classical methods and

T. Özcan • F. Tüysüz (✉)
Department of Industrial Engineering, İstanbul University, 34320, İstanbul, Turkey
e-mail: tuncay.ozcan@istanbul.edu.tr; fatih.tuysuz@istanbul.edu.tr

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_7

modern heuristic methods. Classical methods which are exponential smoothing, regression, Box–Jenkins autoregressive integrated moving average (ARIMA), generalised autoregressive conditionally heteroskedastic (GARCH) methods are based on mathematical and statistical models. Modern heuristic methods use artificial intelligence techniques such as artificial neural networks (ANN) and evolutionary computation. The exponential smoothing, regression and ARIMA methods are linear methods which are based on the Box and Jenkins method. Since these methods rely on long past data, they are generally inappropriate when using limited past data (Xia and Wong 2014). The GARCH method is a nonlinear method and its underlying assumption is that data-generating process of the time series is constant which may limit its applicability. Modern heuristic methods consist of neural networks, fuzzy systems, grey forecasting and hybrid models which are used for nonlinear forecasting models.

The accuracy of forecasting depends on both the characteristics of data and the forecasting model used. The important point is selecting the most appropriate model(s) which is (are) also suitable to the characteristics of the data. It should be considered that each method has some benefits and limitations. For example, regression and ARIMA methods have the limitation of handling nonlinear data. Due to this reason, the best approach can be applying more than one method or model and comparing with each other with respect to the predetermined measures.

Many of the countries face the problem of increasing rates of health care expenditures (HCEs) due to many reasons such as aging population, intensity of healthcare use, change in population structure, technological change (Yu et al. 2015; Blanco-Moreno et al. 2013; Prieto and Lago-Peñas 2012). For the determinants or drivers of HCEs, the interested readers can glance at Astolfi et al. (2012); Prieto and Lago-Peñas et al. (2013); Hartwig and Sturm (2014). According to the OECD Health statictics in 2010, the HCEs to gross domestic product (GDP) ratio for OECD countries reached 11.90% in 2008 which was 8.35% in 1980. According to the OECD health statistics in 2015, health care spending has been growing at around 2.5% per year since 2010. HCE has been increasing constantly and has doubled in the last 50 years. Increased life expectancy combined with declining birth rates causes an ageing population which is one of the main determinants of HCEs. It is also predicted that the ratio of the people who are 60 and over will increase in the world during 2015 and 2080 (Marešová et al. 2015). Although health care demand and HCEs seem to have increased in developed countries, there is no guarantee how it will evolve in both short and long term (Gille and Houy 2014).

In terms of health care management, understanding and predicting HCEs is vital for efficient allocation of resources and effective healthcare policy making, thus taking the necessary action plans. Forecasting in the field of health is quite important for predicting events or situations related to health such as demand for health services and health care needs which facilitates developing the true strategies and minimizing the risks. Forecasting for health requires reliable data and information, as well as appropriate quantitative tools for the prediction (Soyiri and Reidpath 2013).

Due to these reasons, in this study, it is aimed at presenting an analytical approach for forecasting HCEs. Three different strategies are proposed which are rolling mechanism, training data size optimization and parameter optimization to improve forecasting accuracy of grey prediction models. Genetic algorithms (GA) which is one of the most widely used meta-heuristic optimization techniques is applied for data size and parameter optimization of the grey forecasting models. An application of the proposed models for the prediction of HCEs in Turkey and their comparison are also presented.

7.2 Literature Review

7.2.1 Health Care Expenditure Literature

In this section, studies in literature related to the forecasting of health and HCEs which are limited with respect to the studies related to grey forecasting are given. Abdel-Aal and Mangoud (1998) apply two univariate time-series analysis methods for predicting the monthly patient volume at a community medicine primary health care clinic in Saudi Arabia. Models are based on 9 year data and forecasts are made for 2 years.

Getzen (2000) studies short, medium and long term forecasts of health expenditure growth by using the data from 1966 to 1997. The author models both lagged inflation and growth in health employment as control variables in a simple regression model.

Lee and Miller (2002) present a probabilistic model based on stochastic time series to quantify the uncertainty in forecast of HCE. Their model has some advantages over high-low scenario based forecasting approach. They calculate 95% probability intervals as forecasts and indicate great uncertainty about future HCEs.

Munkin and Trivedi (2003) present a simulation-based econometric model for forecasting demand for health care in US. Their model is parametric and Bayesian, and is applied to two different micro data sets.

Wang (2009) proposes and evaluates the usage of equity market information to better forecast the future growth of HCE at a 1-year horizon. US data based results show that industry-specific stock returns have some predictive power for personal HCE and hospital care, durable medical equipment and prescription at a 1-year horizon are its major components.

Ferrand et al. (2011) perform an ARIMA intervention analysis on 6 specific antidepressant drugs and on overall antidepressant spending. Individual ARIMA models are applied for each drug and time series intervention analysis is used for testing the effect of external events on expenditure. Forecasts from the models are also compared with a holdout sample of actual expenditure data.

Van Baal and Wong (2012) investigate the relevance of time-to-death (TTD) according to macrolevel HCE forecasts. They use age-and gender-specific time-series data and show that changes in mortality rates can be related to age-specific per capita HCE.

Eswaran and Logeswaran (2012) present a dual hybrid model which combines linear regression, neural network and fuzzy logic to be used as a decision support system in health care applications.

Astolfi et al. (2012) present a comparative analysis of 25 forecasting models applied for policy analysis in OECD countries used by governments and other related organisations. They classify these models into three groups as micro simulation, component-based, and macro level. They state that no model is superior to the others and the choice of which model to use depends on the policy questions to be addressed.

Maisonneuve and Martins (2013) propose a new set of public health and long-term HCE projections by 2060. Their model combines both demographic and non-demographic drivers. The evolution of each driver is separately predicted and then combined to calculate the future growth of total health spending. A simplified logarithmic growth equation is used to reflect the evolution of expenditure rates.

Huarng and Yu (2014) propose a new quantile information criterion (NQIC) and its thresholds to improve the quantile forecasting performance. They also apply the presented model to forecast HCE of Taiwan and 24 OECD countries.

Aboagye-Sarfo et al. (2015) develop a multivariate vector autoregressive moving average (VARMA) model for forecasting demand for emergency department (ED) in Western Australia. They also compare the performance of the VARMA models with univariate ARMA models and Winters' method. The VARMA model provides better results than the ARMA and Winters' method.

Jalalpour et al. (2015) develop a publicly available toolbox for forecasting demand for health services. It is implemented in MATLAB environment with developed generalized autoregressive moving average (GARMA) models with discrete-valued distributions. They show the validity of the toolbox on two example health care data sets and conclude that GARMA models can outperform the traditional Gaussian models.

The next section presents a detailed literature review for grey forecasting methods and applications.

7.2.2 Grey Forecasting Literature

Grey forecasting is one of the most essential parts in the grey system theory. There are many studies related to grey forecasting and can be summarized as follows. Tseng et al. (2001) suggst that GM (1,1) model is not adequate for forecasting time series with seasonality and propose a hybrid method that combines the GM (1,1) model and the ratio-to-moving-average deseasonalization. They compare the performance of the hybrid model against other four models which are the seasonal time series ARIMA model (SARIMA), the neural network back-propagation model combined with grey relation, the GM (1,1) grey model with raw data, the GM (1,N) grey model combined with grey relation. According to the application results, except

for the out of sample error, the mean squares error (MSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) of the hybrid model are the lowest.

Yao et al. (2003) present an improved grey-based estimation algorithm developed to predict a very-short-term electric power demand. They apply the transformed grey model and the concept of average system slope to improve the prediction.

Zhang et al. (2003) present the unequal interval revised grey model (UIRGM) (1,1) for original series with unequal intervals and sharp variation. The model is applied to estimate element concentration determined by oil spectrometric analysis of diesel engines, and acceptable results are obtained.

Hsu and Chen (2003) propose an improved grey GM (1,1) model that combines residual modification with artificial neural network sign estimation. They apply the proposed model for forecasting Taiwan's power demand and find that their model could improve the prediction accuracy of the original grey model.

Jiang et al. (2004) apply a grey forecasting method to predict the air cooled water chiller units' operating energy performance. They also assess the usability of grey forecasting in fault detection and diagnostics and energy management systems for air conditioning and refrigeration systems.

Wu and Chen (2005) propose an integrated forecasting method which combines GMC (1,N) model with the improved grey relational analysis (GRA). The the convolution technology is integrated in GM (1,N) model to obtain GMC (1,N) model.

Chang et al. (2005) take the P value in univariate grey model as a variable and modify the grey model. They apply their method for forecasting the Taiwan semiconductor industry production and indicate that taking the P value as a variable of time can generate more accurate forecasts.

Zhou et al. (2006) propose a trigonometric grey forecasting approach for predicting electricity demand in which GM (1,1) model is combined with the trigonometric residual modification technique. The proposed approach enables to obtain a reasonable grey prediction interval.

Akay and Atak (2007) propose a grey model with rolling mechanism approach to forecast total and industrial electricity consumption of Turkey. Results show that proposed approach estimates more accurate results than the results of Model of Analysis of the Energy Demand (MAED) used by The Turkish Ministry of Energy and Natural Resources.

Hsu and Wang (2007) evaluate the original and Bayesian grey forecasting models for the integrated circuit industry. Various parameters and efficiency of three forecast models are compared and the Bayesian grey model gives the most accurate results.

Wang and Hsu (2008) propose grey forecasting integrated genetic algorithms approach to predict the output and trends of Taiwan's high technology industry. Genetic algorithms is used to estimate the parameters of the grey model to minimize forecasting error.

Chen (2008) proposes nonlinear grey Bernoulli model (NGBM) by properly choosing power n to predict the unemployment rate of ten selected countries for 2006.

Chen et al. (2008) study the feasibility and effectiveness of a grey model with the concept of the Bernoulli differential equation. The proposed nonlinear grey Bernoulli model (NGBM) is tested with three studies in literature and the results prove the proposed NGBM is feasible and efficient.

Li et al. (2009) use the trend and potency tracking method (TPTM) to reveal the hidden information from data and to build an adaptive grey prediction model, AGM (1,1). The experimental results show that the proposed model could increased the prediction accuracy for small samples.

Hsu (2009) proposes a grey model with factor analysis for the multi-factor forecasting problems. The proposed two multivariable grey forecasting models integrated with genetic algorithm are applied for forecasting the integrated circuit output Taiwan.

Huang and Jane (2009) integrate the moving average autoregressive exogenous (ARX) prediction model with grey systems theory and rough set (RS) theory to develop an automatic stock market forecasting and portfolio selection mechanism. The application result shows that the proposed approach has better forecasting accuracy than the GM (1,1) method.

Lin et al. (2009) introduce a grey forecasting model in which an integration equation is used in calculating background value in order to eliminate the error term. They also integrate Fourier series and exponential smooth technique into the model to reduce the periodic and stochastic residual errors.

Hsu et al. (2009) propose Markov–Fourier grey model (MFGM) to estimate the turning time of Taiwan weighted stock index.

Sheng-qiang et al. (2009) analyze the factors affecting coal and gas outburst. They apply grey correlation method to find the input parameters of artificial neural network (ANN). Then, the neural network model is applied for forecasting coal and gas outburst in China.

Xie and Liu (2009) propose a l discrete grey forecasting model (DGM) and present the optimization models of DGM. The model modifies the algorithm of original GM (1,1) model to improve the tendency catching ability. The results show that the proposed model and its optimized models can increase the prediction accuracy.

Tsaur (2010) proposes a hybrid fuzzy grey regression model for solving crisp-input and fuzzy-input limited time series. The author gives an illustrative example for predicting LCD TV demand by applying the interval grey regression model with limited and interval-input data.

Kumar and Jain (2010) apply three time series models which are grey-Markov model, grey model with rolling mechanism, and singular spectrum analysis (SSA) to forecast the conventional energy consumption of India.

Chen et al. (2010) consider some cases using nonlinear grey Bernoulli model (NGBM) to demonstrate its ability in forecasting nonlinear data. The authors improve NGBM by Nash equilibrium concept to enhance the modeling precision which is supported by an application.

Hsu (2010) proposes an improved nonlinear grey Bernoulli model by using genetic algorithms to determine the optimal parameters for forecasting using

small amount data. The application results show that the improved nonlinear grey Bernoulli model is more accurate and outperforms the traditional GM (1,1) model and grey Verhulst model.

Hsu (2011) proposes an improved transformed grey model based on a genetic algorithm (ITGM (1,1)) and applies it to the output of the opto-electronics industry in Taiwan. The author compares the proposed model with three grey forecasting models and finds that ITGM (1,1) gives better forecasting performance.

Li et al. (2011a) present an improved forecasting model by extending the data transforming approach and generalizing a building procedure for the grey model to grasp the data outline and information trend. The experimental results show that the proposed method improves the forecasting accuracy of grey model.

Li et al. (2011b) propose improved grey dynamic model GM (2,1) which is a second order single variable grey model to improve the forecasting accuracy. They express the derivative and background value of GM (2,1) model as grey number. They also apply the proposed method to short term load forecasting problem and show that it has better performance.

Lee and Tong (2011) develop an improved grey forecasting model, which combines residual modification with genetic programming sign estimation. They show the effectiveness of the proposed forecasting model with a real case of Chinese energy consumption.

Huang et al. (2011) apply a method based on grey system theory and multivariate linear regression for establishing a real estate demand-analyzing model.

Zhan-Li and Jin-Hua (2011) present a forecasting model that features both the GM (1,1) prediction model and the Markov chain prediction model to predict fire accidents. They also analyze a sample and the results show that the proposed Gray-Markov model has a high estimated sensitivity.

Lei and Feng (2012) propose a grey model to improve the performance of traditional grey models for short-term electricity price prediction in competitive power markets. The reference sequence is determined with correlation method and particle swarm optimization algorithm (PSO) is used instead of least square method (LSM) to identify the parameters.

Tang and Yin (2012) apply GM (1,1) and GM (1,1) rolling models for forecasting education expenditure and school enrollments. They compare the results with exponential smoothing technique and show that GM (1,1) rolling model is as good as exponential smoothing.

Li et al. (2012) use adaptive grey forecasting for predicting short term energy consumption. They compare the results with back propagation neural networks (BPN) and support vector regression (SVR).

Ou (2012) presents an improved grey forecasting model by modifying background value calculation and by using genetic algorithm to find the optimal parameters. The proposed model is applied for predicting the agricultural output of Taiwan and the author obtains highly accurate results.

Alvisi and Franchini (2012) propose a forecasting model based on neural networks for river stage. In this model, the parameters of artificial neural network are represented by grey numbers and the output of the model is an interval which better quantifies the imprecision or the uncertainty of the forecasted value.

Kang and Zhao (2012) apply classical grey model and an improved grey model combined with moving average method and Markov model for forecasting long term power load, and show that the accuracy of improved grey model is significantly higher than the ordinary model.

Zhao et al. (2012) develop a high-precision hybrid model based on grey theory and differential evolution (DE) algorithm to estimate the per capita annual net income of rural households in China. DE algorithm is used to optimize λ parameter in GM (1,1) to obtain more accurate results.

Pao et al. (2012) apply nonlinear grey Bernoulli model (NGBM) to forecast CO_2 emissions, energy consumption and real GDP for China, and propose an iterative algorithm for the optimization of parameters in NGBM.

Wang et al. (2013) develop a grey-forecasting interval-parameter mixed-integer programming approach for supporting electric-environmental management at the municipal level in Beijing. A grey forecasting model, GM (1,1), is used to assist in finding variations in electric demand in the near future.

Ma et al. (2013) develop a hybrid grey with rolling mechanism which is optimized by particle swarm optimization algorithm. The model is applied to a real dataset and projection for iron ore imports and total consumption in China for the next 5 years.

Benítez et al. (2013) develop a modified grey forecasting model to predict routes passenger demand growth in the air transportation industry. Modification is done for the GM model to reduce tendency calculations as time grows.

Xie et al. (2013) propose a grey forecasting model based on non-homogeneous index sequence. They show that the models based on homogeneous index sequence are all special cases of nonhomogeneous index sequence.

Cui et al. (2013) propose a grey forecasting model naned NGM model and its optimized model to improve the predictive performance of grey forecasting model. They also develop an algorithm which is based on the least squares method for solving the parameters of the model.

Samvedi and Jain (2013) study the performance of four forecasting methods which are moving average (MA), weighted moving average (WMA), exponential smoothing and grey prediction method during disruptions in a supply chain. The test results show that MA and WMA methods become incompetent during disruptions and are meaningful only during stable times. Grey prediction method becomes the best method, when the frequency of disruption is at least high. They conclude that none of the forecasting methods is the best for every situation.

Liu (2013) applies a grey neural network and input-output combined forecasting model (GNF-IO) to forecast CO_2 emissions volume in China's 42 sectors. Test results show that the model has high simulation and forecasting accuracy on energy consumption.

Hamzacebi and Es (2014) apply optimized grey modeling GM (1,1) in both direct and iterative manners to predict the total electricity energy demand of Turkey for the 2013–2025 period. They find that the direct forecasting approach gives better results than the iterative forecasting approach.

Bahrami et al. (2014) propose a forecasting model which combines the wavelet transform and grey model which is improved by particle swarm optimization (PSO) algorithm for short term electricity load forecasting.

Froelich and Salmeron (2014) address the problem of forecasting multivariate, interval-valued time series and propose to use fuzzy grey cognitive maps. They also apply a genetic algorithm and show that the proposed evolutionary algorithm is applicable for multivariate interval-valued time series.

Jiang et al. (2014) present a hybrid short term demand forecasting model which combines the ensemble empirical mode decomposition (EEMD) and grey support vector machine (GSVM) models. The presented approach is applied to a real case in China and it outperforms ARIMA and support vector machine methods.

Xia and Wong (2014) propose a new seasonal discrete grey forecasting model based on cycle truncation accumulation with amendable items which overcomes seasonality and limited data problems. The proposed model is applied to three sets of real time series and it outperforms the traditional grey forecasting method.

Zeng et al. (2015) propose a method called neural network triangular fuzzy grey forecasting model (NNTFGM (1,1)). In their model, the original GM (1,1) model is improved by the triangular fuzzy (TF) series and then neural networks are used for forecasting of fluctuating TF series.

Xie et al. (2014) use grey forecasting model for rating prediction in recommender systems and perform experiments on two public movie data sets. The results show that gray prediction models can overcome the problem of data sparsity, benefit from data correlations, and are better than traditional memory based methods.

Chang et al. (2015) present a box plot to analyze data features and propose a new algorithm for the background values in the grey model to improve the forecasting.

Intharathirat et al. (2015) apply optimized multivariate grey model GM (1,5) for forecasting the municipal solid waste quantity in Thailand. They use correlation analysis and grey relational analysis (GRA) to identify and classify the influencing factors affecting the waste collected.

Xie et al. (2015) apply a model for forecasting energy production and consumption of China between 2006 and 2011. They use an optimized GM (1,1) model to forecast the total energy production and consumption and also a Markov model based on quadratic programming to forecast the trends of energy production and consumption structures.

Wu et al. (2015) use a multi-variable grey model to examine the relationship among energy consumption, urban population, economic growth and CO_2 emissions in the BRICS (Brazil, Russia, India, China, and South Africa) countries betwen 2004 and 2010. The CO_2 emissions of the countries are estimated by the grey model with rolling mechanism considering energy consumption and economic growth.

Sun et al. (2016) develop a model for predicting the variation in tourism demand. Their model combines a Cuckoo search optimization algorithm and an optimal input subset for the GM (1,1) model. Markov-chain residual modified model is also used to decrease the impact of the randomness and fluctuation whereas Cuckoo search optimization algorithm is used in whitening the grey interval.

Table 7.1 presents a summary of the grey forecasting literature based on the approach used.

Based on the reviewed studies, a basic comparison of the grey forecasting models which can be helpful for the interested readers is given in Table 7.2.

Although GM (1,1), NGBM (1,1) and GM (1,N) are the three main grey forecasting models, the literature review indicates that the recent studies mainly focus on modified or improved forms of these models which aim at increasing the accuracy of forecasting. These developed models are usually in the form of applying some optimization techniques for the parameters used in the model. In this study, the same approach will be handled and presented in detail.

7.3 Grey Forecasting

Grey systems theory which was developed by Deng (1982) can be used in solving uncertainty problems which contain discrete data and incomplete information. The most important advantages of the grey system theory are that it can achieve successful results in relatively small quantities in case of uncertainty and it helps to analyze and model systems with limited or incomplete information.

The grey system theory contains five main parts: grey forecasting, grey relational analysis, grey decision making, grey programming and grey control (Wei 2011). Grey forecasting is one of the most essential parts in the grey theory.

Compared with traditional statistical prediction models, the advantages of grey prediction models can be summarized as follows (Zhou et al. 2006; Feng et al. 2012; Ren et al. 2012; Tsai 2016):

- When the number of data is not enough for traditional statistical methods, grey prediction models can be used. In the grey prediction models, only small amounts of data are required to describe the system behavior.
- The distribution of sample populations is not required.
- The noise of the original data is reduced by using accumulated generating operation (AGO).
- Grey prediction models involve simple calculations.

The data required for grey forecxasting can be easily obtained and the data collection time and costs are considerably lower than those required for other methods.

Before presenting the models applied for the application, we will briefly present the basic grey forecasting models, that are GM (1,1) and NGBM (1,1).

Table 7.1 Summary of the grey forecasting literature

Study	Approach (grey model type)	Additional method/technique	Application area
Jiang et al. (2004)	GM (1,1)	–	Operating energy performance forecasting of ACWC units
Samvedi and Jain (2013)	GM (1,1)	–	Theoretical example for comparison
Xie et al. (2014)	GM (1,1)	–	Rating prediction in recommender systems
Akay and Atak (2007)	GM (1,1) with rolling mechanism	–	Turkey's total and industrial electricity consumption forecasting
Tang and Yin (2012)	GM (1,1) and GM (1,1) with rolling mechanism	–	Forecasting education expenditure and school enrollments
Yao et al. (2003)	Improved GM (1,1)	Average system slope technique	Very-short-term electric power demand forecasting
Zhang et al. (2003)	Improved GM (1,1)	Unequal intervals	Marine diesel engines' wear trend forecasting
Hsu and Chen (2003)	Improved GM (1,1)	Residual modification with ANN estimation	Power demand forecasting of Taiwan
Zhou et al. (2006)	Improved GM (1,1)	Trigonometric residual modification technique	Electricity demand forecasting
Wang and Hsu (2008)	Improved GM (1,1)	Genetic algorithms	Forecasting the output of high technology industry in Taiwan
Li et al. (2009)	Improved GM (1,1)	Trend and potency tracking method	Early manufacturing data forecasting
Lee and Tong (2011)	Improved GM(1,1)	Genetic programming	Chinese energy consumption forecasting
Hsu (2011)	Improved GM (1,1)	Genetic algorithms	Forecasting output of the opto-electronics industry in Taiwan
Li et al. (2011a)	Improved GM (1,1)	Inverse accumulating generation operator	Titanium alloy fatigue limit forecasting

(continued)

Table 7.1 (continued)

Study	Approach (grey model type)	Additional method/technique	Application area
Lei and Feng (2012)	Improved GM (1,1)	Correlation method and particle swarm optimization algorithm	Short-term electricity price prediction
Ou (2012)	Improved GM (1,1)	Genetic algorithms	Forecasting agricultural output of Taiwan
Kang and Zhao (2012)	Improved GM(1,1)	Moving average method and Markov model	Forecasting long term power load
Xie et al. (2013)	Improved GM (1,1)	Non-homogeneous index sequence	Theoretical example for comparison
Wang et al. (2013)	Improved GM (1,1)	Mixed-integer programming	Energy forecasting in Beijing, China
Hamzacebi and Es (2014)	Improved GM (1,1)	A numerical optimization algorithm	Forecasting the total electric energy demand of Turkey
Bahrami et al. (2014)	Improved GM (1,1)	Wavelet transform and particle swarm optimization algorithm	Short term electricity load forecasting for Iran and New York
Xia and Wong (2014)	Improved GM (1,1) (seasonal discrete)	Cycle truncation accumulation algorithm	Fashion retailing sales forecasting
Zeng et al. (2015)	Improved GM (1,1)	Fuzzy time series and neural network method	Forecasting consumer price indices and the power load of one district of China
Chang et al. (2015)	Improved GM (1,1)	A new formula for background value calculation	Forecasting variations in manufacturing process
Xie et al. (2015)	Improved GM (1,1)	Markov approach based on quadratic programming	Forecasting China's energy production and consumption
Sun et al. (2016)	Improved GM (1,1)	Markov chain and cuckoo search optimization	Forecasting the annual foreign tourist arrivals to China
Chang et al. (2005)	Modified GM (1,1)	–	Taiwan semiconductor industry production forecasting
Xie and Liu (2009)	Modified GM (1,1)	Recursive equations for parameters	Theoretical examples for comparison
Benítez et al. (2013)	Modified GM (1,1)	–	Forecasting routes passenger demand growth in the air transportation

Cui et al. (2013)	Modified GM (1,1)	–	Theoretical example for comparison
Tseng et al. (2001)	Hybrid GM (1,1)	Ratio-to-moving-average deseasonaliztion method	Taiwan's machinery industry total production forecasting
Huang and Jane (2009)	Hybrid GM (1,1)	The moving average autoregressive exogenous method and rough sets	Stock market forecasting
Lin et al. (2009)	Hybrid GM (1,1)	Fourier time series analysis and exponential smoothing techniques	Material stock index forecasting
Zhao et al. (2012)	Hybrid GM (1,1)	Differential evolution algorithm	Forecasting the per capita annual net income of rural households in China
Liu (2013)	Hybrid GM (1,1)	Neural network	Forecasting CO_2 emissions volume in China's 42 sectors
Ma et al. (2013)	Hybrid GM (1,1) with rolling mechanism	Particle swarm optimization algorithm	Forecasting iron ore imports and total consumption in China
Hsu and Wang (2007)	Bayesian GM (1,1)	Markov Chain Monte Carlo technique	Forecasting the output of integrated circuit industry in Taiwan
Zhan-Li and Jin-Hua (2011)	Markov model GM (1,1)	Markov chain	Forecasting fire accidents
Kumar and Jain (2010)	Markov model GM (1,1)	Markov chain	Forecasting energy consumption in India
Hsu et al. (2009)	Markov–Fourier grey model, MFGM (1,1)	Fourier series and Markov state transition	Forecasting the turning time of Taiwan weighted stock index
Chen (2008)	Nonlinear grey Bernoulli, NGBM (1,1)	Bernoulli differential equation	Forecasting the annual unemployment rate of 10 selected countries
Chen et al. (2008)	Nonlinear grey Bernoulli, NGBM (1,1)	Bernoulli differential equation	Forecasting foreign exchange rates of Taiwan's major trading partners
Chen et al. (2010)	Nonlinear grey Bernoulli, NGBM (1,1)	Nash equilibrium	Forecasting the monthly Taiwan stock indices

(continued)

Table 7.1 (continued)

Study	Approach (grey model type)	Additional method/technique	Application area
Pao et al. (2012)	Nonlinear grey Bernoulli, NGBM (1,1)	A numerical iterative method	Forecasting CO_2 emissions, energy consumption and real GDP for China
Hsu (2010)	Improved nonlinear grey Bernoulli, NGBM (1,1)	Genetic algorithms	Forecasting Taiwan's integrated circuit industry output
Tsaur (2010)	Fuzzy GM (1,1)	Fuzzy grey regression	Forecasting LCD TV demand
Li et al. (2011b)	Improved grey dynamic model GM (2, 1)	–	Short term load forecasting
Wu et al. (2015)	GM (1,N) with rolling mechanism	–	Dorecasting the CO2 emissions in BRICS countries
Wu and Chen (2005)	Improved GM (1,N)	Grey relational analysis	Taiwan's internet access population forecasting
Hsu (2009)	Improved GM (1,N)	Factor analysis and genetic algorithms	Taiwanese integrated circuit output forecasting
Intharathirat et al. (2015)	Improved GM (1,N)	Correlation analysis and grey relational analysis	Forecasting the municipal solid waste quantity in Thailand
Huang et al. (2011)	Grey multivariate linear regression	–	Real estate market demand forecasting
Li et al. (2012)	Adaptive grey prediction model	–	Short term energy forecasting
Alvisi and Franchini (2012)	Grey numbered artificial neural network	–	River stage forecasting
Froelich and Salmeron (2014)	Fuzzy grey cognitive maps	Genetic algorithms	Experiments with real meteorological data for comparison
Jiang et al. (2014)	Hybrid grey support vector machine	–	Short-term forecasting of high-speed rail demand in China

Table 7.2 A basic comparison of the grey forecasting models

	GM (1,1)	NGBM (1,1)	GM (1,N)
Relationship type	Linear	Nonlinear	Linear
# of dependent variables	1	1	1
# of independent variables	1	1	N
Data requirement	Low (at least 4)	Low (at least 4)	Low (at least 4)
# of parameters used	1	2	N + 1
Calculation complexity	Low	Medium	High

7.3.1 GM (1,1) Model

The GM (1,1) model which provides good prediction by using limited data, is the basic grey forecasting model. GM (1,1) indicates one variable and one order grey forecasting model (Wang and Hsu 2008).

Let $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(i), \ldots, x^{(0)}(n)\}$ be an original time series data. Here, $x^{0}(i)$ corresponds to the time series data at time i and n is the sample size of time series data. The calculation procedure of GM (1,1) model is described as follows (Tseng et al. 2001; Cui et al. 2013):

Step 1: A new sequence $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(i), \ldots, x^{(1)}(n)\}$ is constructed using accumulated generating operation (AGO) technique by using Eq. (7.1).

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), \quad k = 1, 2, \ldots, n \tag{7.1}$$

Step 2: A sequence $Z^{(1)} = \{z^{(1)}(2), \ldots, z^{(1)}(i), \ldots, z^{(1)}(n)\}$ is obtained by consecutive values of sequence $X^{(1)}$ using Eq. (7.2).

$$z^{(1)}(k) = \alpha x^{(1)}(k) + (1-\alpha)\, x^{(1)}(k-1), \qquad \forall k = 2, 3, \ldots, n, \quad 0 \le \alpha \le 1 \tag{7.2}$$

α is a production coefficient of the background value. Parameter α is taken as 0.5 in most problems.

Step 3: The first order–one variable grey differential equation of GM (1,1) is established as follows.

$$x^{(0)}(k) + az^{(1)}(k) = b, \quad k = 2, 3, \ldots, n \tag{7.3}$$

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b, \quad k = 2, 3, \ldots, n \tag{7.4}$$

In these equations, a is developing coefficient and b is grey input. The optimized values of parameter a and b can be calculated with Eq. (7.5) using the least square method.

$$[a,b]^T = \left(B^T B\right)^{-1} B^T Y \tag{7.5}$$

where

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{7.6}$$

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ -z^{(1)}(4) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \tag{7.7}$$

Step 4: The predicted values of the accumulated sequence are obtained by using Eq. (7.8).

$$\widehat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right) e^{-ak} + \frac{b}{a}, \quad k = 1, 2, \ldots, n \tag{7.8}$$

Step 5: The predicted values of the original sequence are calculated with Eq. (7.9) by using the inverse accumulated generating operation (IAGO).

$$\widehat{x}^{(0)}(k+1) = \widehat{x}^{(1)}(k+1) - \widehat{x}^{(1)}(k), \quad k = 1, 2, \ldots, n \tag{7.9}$$

7.3.2 Nonlinear Grey Bernoulli Model

GM (1,1) model always generates exponentially increasing and decreasing series. GM (1,1) model is inefficient to forecast time series with randomness, nonlinearity and fluctuation. In order to overcome these deficiencies, nonlinear grey Bernoulli model – NGBM (1,1) – can be used.

Nonlinear grey Bernoulli model (NGBM (1,1)) is constructed by combining both the classical GM (1,1) model and the Bernoulli equation (Hsu 2010). In the NGBM (1,1), the Step 1 and Step 2 are the same as in GM (1,1) model. The other steps of NGBM (1,1) model are described as follows (Zhou et al. 2009; Chen et al. 2010; Hsu 2010).

Step 3: The first order Bernoulli differential equation of NGBM (1,1) is established as follows.

$$x^{(0)}(k) + az^{(1)}(k) = b\left[z^{(1)}(k)\right]^{\gamma}, \quad k = 2, 3, \ldots, n \tag{7.10}$$

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b\left[z^{(1)}(k)\right]^{\gamma}, \quad k = 2, 3, \ldots, n \tag{7.11}$$

where γ belongs to any real number except 1. As mentioned above, a is developing coefficient and b is grey input. The optimized values of parameter a and b can be estimated with Eq. (7.12) by the least-square method.

$$[a, b]^T = \left(B^T B\right)^{-1} B^T Y \tag{7.12}$$

where

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{7.13}$$

$$B = \begin{bmatrix} -z^{(1)}(2) & \left[z^{(1)}(2)\right]^{\gamma} \\ -z^{(1)}(3) & \left[z^{(1)}(3)\right]^{\gamma} \\ -z^{(1)}(4) & \left[z^{(1)}(4)\right]^{\gamma} \\ \vdots & \vdots \\ -z^{(1)}(n) & \left[z^{(1)}(n)\right]^{\gamma} \end{bmatrix} \tag{7.14}$$

Step 4: The predicted values of the accumulated sequence are obtained by using Eq. (7.15).

$$\widehat{x}^{(1)}(k+1) = \left[\left(x^{(0)}(1)^{(1-\gamma)} - \frac{b}{a}\right) e^{-a(1-\gamma)k} + \frac{b}{a}\right]^{1/(1-\gamma)}, \quad \gamma \neq 1, k = 1, 2, \ldots, n \tag{7.15}$$

In the above equation, when power index (γ) is 0, Eq. (7.15) converts to Eq. (7.8). Therefore, GM (1,1) model is a special case of NGBM (1,1) model. Similarly, when $\gamma = 2$, the NGBM (1,1) model converts to Grey-Verhulst model (Chen 2008; Hsu 2010). NGBM (1,1) model has greater flexibility than GM (1,1) and Grey-Verhulst models with adjustable parameter γ (Zhou et al. 2009).

Step 5: The predicted values of the original sequence are calculated with Eq. (7.16) by using the inverse accumulated generating operation (IAGO).

Table 7.3 HCE per capita in Turkey from 1995 to 2013

Years	HCE per capita ($)	Years	HCE per capita ($)
1995	93,234	2005	367,505
1996	113,575	2006	424,418
1997	125,988	2007	560,617
1998	149,685	2008	628,503
1999	180,927	2009	522,434
2000	197,222	2010	565,603
2001	148,970	2011	555,609
2002	180,946	2012	568,968
2003	231,853	2013	607,708
2004	297,903		

$$\widehat{x}^{(0)}(k+1) = \widehat{x}^{(1)}(k+1) - \widehat{x}^{(1)}(k), \quad k = 1, 2, \ldots, n \tag{7.16}$$

Before giving the methodology and the proposed models, the data set will be given in the next section in order to make the methodology more easily be tracked by the readers.

7.3.3 Data Set

This study aims to predict the health care expenditure (HCE) per capita by using grey forecasting models for the future years in Turkey. To this end, the annual data of HCE per capita ($) of Turkey for the period 1995–2013 are provided by World Health Organization Global Health Expenditure database, as shown in Table 7.3 and Fig. 7.1 (see http://apps.who.int/nha/database).

The data set of Turkey's HCE includes annual data from 1995 to 2013 ($n = 19$). According to this data set, HCE in the Turkey varies substantially over time. Health care spending per capita increased by approximately 6.5-fold from 1995 to 2013. At the same time, as a share of Gross Domestic Product (GDP), HCE in Turkey grew at roughly 65% between 1995 and 2013, increasing from 3.37% in 1995 to 5.59% in 2013.

7.4 Methodology

In this study, three different strategies are proposed that are rolling mechanism, training data size optimization and parameter optimization to improve the forecasting accuracy of grey prediction models. These strategies can be summarized as follows.

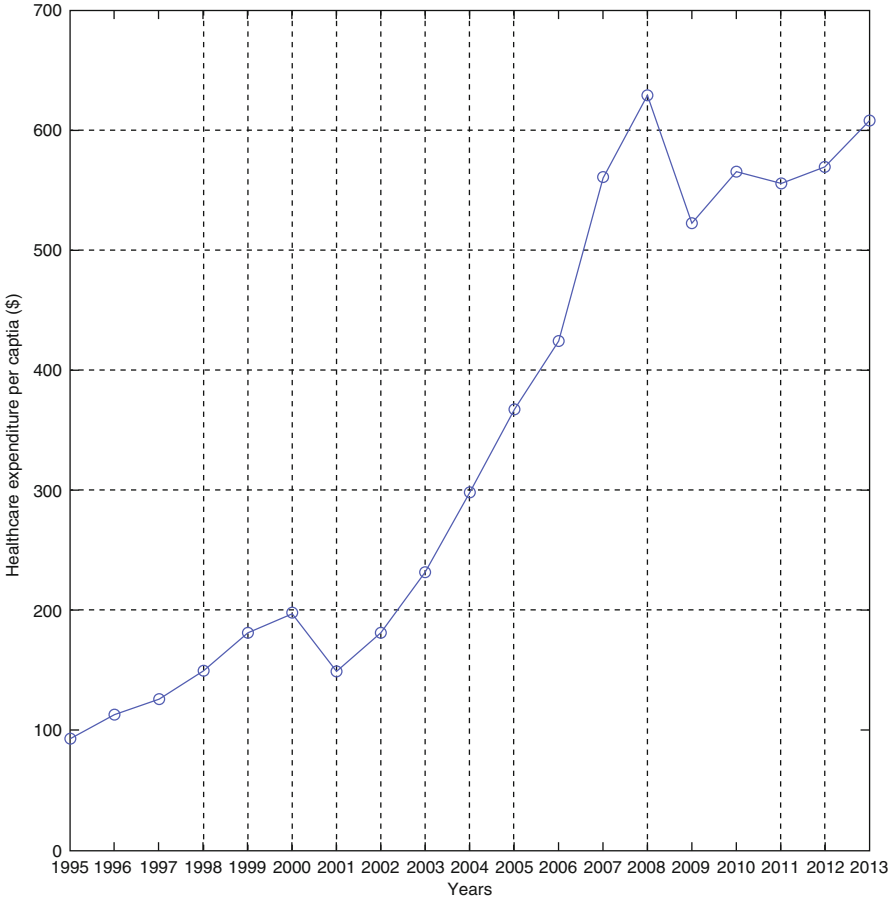

Fig. 7.1 HCE per capita in Turkey from 1995 to 2013

Rolling mechanism is an effective technique that is used to construct the forecasting model. This mechanism can be explained as follows:

Suppose that n denotes the size of time series data. The data set can be divided into training data set and testing data set. Let t denote the size of training data set ($t<n$) and l denote the size of testing data set ($l<n$). In the rolling mechanism, for example; the predicted value for data point k ($\widehat{x_0}(k)$) is calculated using recent t data ($x_0(k-t), x_0(k-t+1), \ldots, x_0(k-1)$). After the predicted value is obtained, the oldest data ($x_0(k-t)$) is removed from training data set, and the newly data ($x_0(k)$) is added to the end of the training data set. In the next iteration, $x_0(k-t+1), x_0(k-t+2), \ldots, x_0(k)$ data set is used to predict the next value $x_0(k+1)$. This procedure is repeated until the size of the testing data.

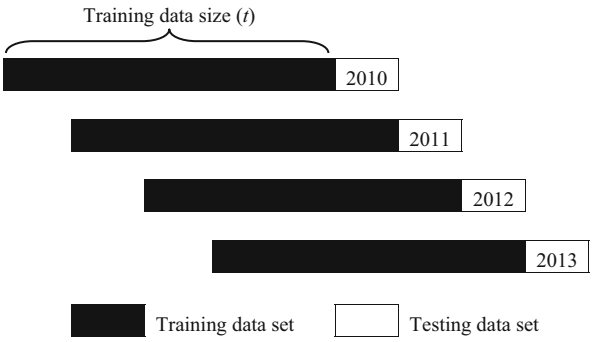

Fig. 7.2 The diagram of the rolling mechanism

In this study, the data from 2010 to 2013 are used as testing data set ($l = 4$). This data set is used to evaluate the performance of the proposed models. Rolling mechanism that is employed to generate the proposed models is given in Fig. 7.2.

Grey prediction models with rolling mechanism are recommended to use only recent data to increase accuracy in future prediction (Wu et al. 2013). At this point, an important problem is to decide how much historical data is optimal in order to obtain an accurate forecast. The size of training data (t) has a significant impact on forecasting accuracy in grey prediction models with rolling mechanism.

The minimum number of data must be four in grey models (Deng 1986). In other words, the minimum size of training data is 4. In this study, the whole data set consists of 19 years ($n = 19$) and the data from 2010 to 2013 are used as testing data set ($l = 4$). Therefore, the maximum size of training data is 15. The optimum training data size (t) is between 4 and 15.

Grey forecasting models contain the adjustable parameters such as production coefficient of the background value (α) and power index (γ). In the traditional grey models, the values of these parameters are fixed. These fixed values are not optimal for all data sets. Therefore, the parameter optimization is an important issue for improving the prediction accuracy of these models. In the literature, meta-heuristic algorithms such as genetic algorithms (Wang and Hsu 2008; Hsu 2010; Lee and Tong 2011; Ou 2012) and particle swarm optimization (Zhou et al. 2009; Bahrami et al. 2014) were successfully applied to estimate the parameters of grey forecasting models. In this study, genetic algorithm is used to optimize the parameters of these models and the size of training data.

Genetic algorithm (GA) is a population based stochastic evolutionary computation algorithm and was introduced by Holland (1975). This algorithm is a widely used method in solving combinatorial optimization problems. This algorithm is inspired by natural evaluation.

The basic GA can be expressed as the following:

Step 1: Initialization – Generation of initial solution using random number operator
Step 2: Evaluation – Calculation of fitness value of each individual
Step 3: Selection – Selection of parents for reproduction.

Step 4: Reproduction – Creation of new generation by applying genetic operators such as elitism, crossover and mutation

Step 5: Control – Repetition of Step 2–4 until the stopping criterion is satisfied.

In this study, the mean absolute percentage error (MAPE) is used for comparing the forecasting accuracy of the proposed models. At the same time, the fitness (or objective) function is set by minimizing the MAPE in the training size and parameter optimization problem. MAPE is defined as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{\widehat{x}_0(k) - x_0(k)}{x_0(k)}\right| \times 100\% \tag{7.17}$$

where $x_0(k)$ is the actual value, $\widehat{x}_0(k)$ is the predicted value and n is the number of prediction data.

7.4.1 Healthcare Expenditure Forecasting with Proposed Models

In this study, eight different models are proposed based on GM (1,1), Grey-Verhulst and NGBM (1,1) for solving HCE forecasting problem. In order to compare the forecasting performance of the grey forecasting based models with statistical methods, another model based on ARIMA(p,d,q) is used. These forecasting models can be briefly described as follows:

Model 1: Original GM (1,1) model

In Model 1, the whole data set is used for prediction. For example, the data from 1995 to 2010 are used to generate a model for predicting the value of 2011, while the 2010 value is calculated using the data from 1995 to 2009, and so on. Production coefficient of the background value (α) in this model is fixed at 0.5. In this model, for $k = 16, 17, 18, 19$ (2010–2013 years), the HCE values are predicted to be 715,638; 754,897; 770,873 and 782,013, respectively. The MAPE of the traditional GM (1,1) model is 31.64%.

Model 2: GM (1,1) model with rolling mechanism and training data size optimization

In Model 2, the rolling mechanism is used to obtain the predicted values. The parameter values of GM (1,1) model are fixed. Production coefficient of the background value (α) in this model is taken as 0.5. The problem of training data size optimization can be formulated as the following optimization problem where the decision variable is t and the objective function is to minimize MAPE:

$$\min \quad Z = \frac{1}{4} \sum_{k=16}^{19} \left| \frac{\widehat{x}_0(k) - x_0(k)}{x_0(k)} \right| \times 100\% \tag{7.18}$$

s.t.

$$4 \leq t \leq 15 \tag{7.19}$$

$$t \in \mathrm{N} \tag{7.20}$$

$$\alpha = 0.5 \tag{7.21}$$

We obtain the minimum MAPE (5.65%) and optimal parameter value $t = 4$ by using GA. In this model, for $k = 16, 17, 18, 19$ (2010–2013 years), the HCE values are predicted to be 535,132; 509,371; 581,306 and 566,791, respectively.

Model 3: GM (1,1) model with rolling mechanism and parameter optimization

In Model 3, the rolling mechanism is used to construct the model. The size of the training data is fixed as 15. The problem of parameter optimization can be formulated as the following optimization problem where the decision variable is production coefficient of the background value (α) and the objective function is to minimize MAPE:

$$\min \quad Z = \frac{1}{4} \sum_{k=16}^{19} \left| \frac{\widehat{x}_0(k) - x_0(k)}{x_0(k)} \right| \times 100\% \tag{7.22}$$

s.t.

$$0 \leq \alpha \leq 1 \tag{7.23}$$

$$t = 15 \tag{7.24}$$

We obtain the minimum MAPE (3.52%) and optimal parameter value $\alpha = 0.59$. In this model, for $k = 16, 17, 18, 19$ (2010–2013 years), the HCE values are predicted to be 513,825; 562,132; 588,472 and 605,628, respectively.

Model 4: GM (1,1) model with rolling mechanism, and training data size and parameter optimization

In Model 4, the rolling mechanism is used to obtain the prediction values. The parameter values of GM (1,1) model and the size of the training data are optimized. This optimization problem can be formulated where the decision variables are production coefficient of the background value (α) and the size of the training data (t) and the objective function is to minimize MAPE:

$$\min \quad Z = \frac{1}{4} \sum_{k=16}^{19} \left| \frac{\widehat{x}_0(k) - x_0(k)}{x_0(k)} \right| \times 100\% \tag{7.25}$$

s.t.

$$4 \leq t \leq 15 \tag{7.26}$$

$$t \in N \tag{7.27}$$

$$0 \leq \alpha \leq 1 \tag{7.28}$$

We obtain the minimum MAPE (2.70%) and optimal parameter values ($t = 4$, $\alpha = 0.76$) by using GA. In this model, for $k = 16, 17, 18, 19$ (2010–2013 years), the HCE values are predicted as 560,207; 555,231; 554,064 and 564,212 respectively.

Model 5: Grey-Verhulst model

In Model 5, the whole data set is used for prediction. In this model, production coefficient of the background value (α) and power index (γ) are fixed at 0.5 and 2, respectively. In this model, for $k = 16, 17, 18, 19$ (2010–2013 years), the HCE values are predicted to be 458,302; 519,662; 556,236 and 576,690, respectively. The MAPE of the Grey-Verhulst model is 8.20%.

Model 6: Grey-Verhulst model with rolling mechanism and training data size optimization

In Model 6, the power index (γ) parameter of NGBM (1,1) model is fixed as 2. Production coefficient of the background value (α) in this model is taken as 0.5. The problem of training data size optimization can be formulated as the following optimization problem where the decision variable is t and the objective function is to minimize MAPE:

$$\min \quad Z = \frac{1}{4} \sum_{k=16}^{19} \left| \frac{\widehat{x}_0(k) - x_0(k)}{x_0(k)} \right| \times 100\% \tag{7.29}$$

s.t.

$$4 \leq t \leq 15 \tag{7.30}$$

$$t \in N \tag{7.31}$$

$$\alpha = 0.5 \tag{7.32}$$

$$\gamma = 2 \tag{7.33}$$

We obtain the minimum MAPE (4.93%) and optimal parameter value $t = 12$ by using GA. In this model, for $k = 16, 17, 18, 19$ (2010–2013 years), the HCE values are predicted to be 564,969; 555,217; 504,413 and 557,983, respectively.

Model 7: NGBM (1,1) model with rolling mechanism and parameter optimization

NGBM (1,1) model has greater flexibility than GM (1,1) and Grey-Verhulst models. In Model 7, the rolling mechanism is used to construct the NGBM (1,1) model. The size of the training data is fixed as 15. The problem of parameter optimization can be formulated as the following optimization problem where the decision variables are production coefficient of the background value (α) and power index (γ), and the objective function is to minimize MAPE:

$$\min \quad Z = \frac{1}{4}\sum_{k=16}^{19}\left|\frac{\widehat{x}_0(k) - x_0(k)}{x_0(k)}\right| \times 100\% \tag{7.34}$$

s.t.

$$0 \leq \alpha \leq 1 \tag{7.35}$$

$$\gamma \neq 1 \tag{7.36}$$

$$t = 15 \tag{7.37}$$

We obtain the minimum MAPE (2.23%) and optimal parameter values ($\alpha = 0.54$ and $\gamma = -1.03$). In this model, for $k = 16, 17, 18, 19$ (2010–2013 years), the HCE values are predicted to be 562,748; 572,266; 569,176 and 574,916, respectively.

Model 8: NGBM (1,1) model with rolling mechanism, and training data size and parameter optimization

In Model 8, the parameter values of NGBM (1,1) model and the size of the training data are optimized. This optimization problem can be formulated where the decision variables are production coefficient of the background value (α), power index (γ) and the size of the training data (t), and the objective function is to minimize MAPE:

$$\min \quad Z = \frac{1}{4}\sum_{k=16}^{19}\left|\frac{\widehat{x}_0(k) - x_0(k)}{x_0(k)}\right| \times 100\% \tag{7.38}$$

s.t.

$$4 \leq t \leq 15 \tag{7.39}$$

$$t \in N \tag{7.40}$$

$$0 \leq \alpha \leq 1 \tag{7.41}$$

$$\gamma \neq 1 \tag{7.42}$$

We obtain the minimum MAPE (0.73%) and optimal parameter values ($t = 4$, $\alpha = 0.76$) by using GA. In this model, for $k = 16, 17, 18, 19$ (2010–2013 years), the HCE values are predicted to be 566,473; 556,388; 568,780 and 591,850, respectively.

We used only 19 HCE data in this study. Since the historical data about HCE are limited and are not normally distributed, grey prediction models are suitable for this kind of data set. On the other hand, HCE data have the characteristics of nonlinear nature and random fluctuations. As it can be seen from Model 1, original GM (1,1) model could not fit to this data and its prediction inaccuracy is higher. At this point, the forecasting qualityof GM (1,1) model is improved by using NGBM (1,1) model for nonlinear data sets. Besides, the prediction error of NGBM (1,1) model can be significantly reduced with rolling mechanism, training data size optimization and parameter optimization as in Model 8. This model has lower prediction error than the other seven grey models. Therefore, the Model 8 is further utilized to predict Turkey's HCE.

The optimal and fixed parameter values for the eight different grey models are shown in Table 7.4.

Model 9: ARIMA (*p,d,q*) model with rolling mechanism, and training data size and parameter optimization

In this study, in order to compare the forecasting performance of the grey forecasting based models with statistical methods, optimized ARIMA (*p,d,q*) model is proposed. The ARIMA model is widely used in the areas of non-stationary time series forecasting. Let a time series $\{X_t | t = 1, 2, ..., k\}$ be generated by ARIMA

Table 7.4 The parameter values of the proposed grey models

Model	Coefficient of the background value (α)	Power index (γ)	Training data size (*t*)	Rolling mechanism
Model 1	0.5 (fixed)	0 (fixed)	All data	No
Model 2	0.5 (fixed)	0 (fixed)	4 (optimized)	Yes
Model 3	0.59 (optimized)	0 (fixed)	15 (fixed)	Yes
Model 4	0.76 (optimized)	0 (fixed)	4 (optimized)	Yes
Model 5	0.5 (fixed)	2 (fixed)	All data	No
Model 6	0.5 (fixed)	2 (fixed)	12 (optimized)	Yes
Model 7	0.54 (optimized)	−1.03 (optimized)	15 (fixed)	Yes
Model 8	0.58 (optimized)	−0.87 (optimized)	10 (optimized)	Yes

(p,d,q) process of Box and Jenkins time series model. The ARIMA model can be defined as follows (Box and Jenkins 1976):

$$\phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t \tag{7.43}$$

where d is the number of regular differences, B is a backward shift operator defined as in Eqs. (7.44–7.45).

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \tag{7.44}$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \tag{7.45}$$

where p is the order of the autoregressive component and q is the order of the moving averages component.

In Model 9, the parameters (p,d,q) of original ARIMA model and the size of the training data (t) are optimized. This optimization problem can be formulated as follows:

$$\min \quad Z = \frac{1}{4}\sum_{k=16}^{19}\left|\frac{\widehat{x}_0(k) - x_0(k)}{x_0(k)}\right| \times 100\% \tag{7.46}$$

s.t.

$$0 \le p \le 7 \tag{7.47}$$

$$0 \le d \le 7 \tag{7.48}$$

$$0 \le q \le 7 \tag{7.49}$$

$$4 \le t \le 15 \tag{7.50}$$

We obtain the minimum MAPE (3.32%) and optimal parameter values ($p = 0$, $d = 1$, $q = 0$ and $t = 12$) using GA. In this model, for $k = 16, 17, 18, 19$ (2010–2013 years), the HCE values are predicted to be 547,284; 591,248; 579,502 and 596,968, respectively.

Proposed forecasting models are programmed on MATLAB 2014a. The training data size and/or parameter optimization problems are solved with MATLAB GA Toolbox.

The values of the GA parameters such as the population size (n), the crossover probability (p_c), mutation probability (p_m) and maximum number of generations (G) have a significant impact on the solution quality and computational time of the

algorithm. Therefore, in this study, the robustness of this algorithm is examined for different GA parameter values. At this point, several tests are conducted with values for $n = 10, 20, 30$ and $G = 20, 50, 100$. Additionally, for parameter tuning, the crossover probability is varied in the range [0.5, 1.0] with an increment of 0.1 and the mutation probability is changed in the range [0.01, 0.1] with an increment of 0.01. The default values of the MATLAB optimization toolbox are used for the other parameters such as stopping criteria, elite count and selection function.

The proposed forecasting models are solved with all parameter sets. For each case, 10 independent replications are carried out. MAPE converges fast to a stationary point for each parameter combination. Due to the small size of the problem, the same solutions are easily obtained using different set of parameters for each model. Although the MAPE values do not change for different values of the GA parameters, the number of generations required to reach the stable MAPE values does slightly change. For example, among the all sets of the parameters, $p_c = 0.9$, $p_m = 0.01$, $n = 10$ and $G = 20$ values show the best performance in terms of computational time for Model 8.

The prediction results and MAPE values obtained by the proposed models are given in Table 7.5.

Table 7.5 MAPE and predicted values of the proposed forecasting models

		Model 1		Model 2		Model 3	
Year	Actual value	Predicted value	Error (%)	Predicted value	Error (%)	Predicted value	Error
2010	565,603	715,638	26.53	535,132	5.39	513,825	9.15
2011	555,609	754,897	35.87	509,371	8.32	562,132	1.17
2012	568,968	770,873	35.49	581,306	2.17	588,472	3.43
2013	607,708	782,013	28.68	566,791	6.73	605,628	0.34
MAPE (%) 2010–2013			**31.64**		**5.65**		**3.52**
		Model 4		Model 5		Model 6	
Year	Actual value	Predicted value	Error (%)	Predicted value	Error (%)	Predicted value	Error (%)
2010	565,603	560,207	0.95	458,302	18.97	564,967	0.11
2011	555,609	555,231	0.07	519,662	6.47	555,217	0.07
2012	568,968	554,064	2.62	556,236	2.24	504,413	11.35
2013	607,708	564,212	7.16	576,690	5.10	557,983	8.18
MAPE (%) 2010–2013			**2.70**		**8.20**		**4.93**
		Model 7		Model 8		Model 9	
Year	Actual value	Predicted value	Error (%)	Predicted value	Error (%)	Predicted value	Error
2010	565,603	562,748	0.50	566,473	0.15	547,284	3.24
2011	555,609	572,267	3.00	556,388	0.14	591,248	6.41
2012	568,968	569,176	0.04	568,780	0.03	579.502	1.85
2013	607,708	574,916	5.40	591,850	2.61	596,968	1.77
MAPE (%) 2010–2013			**2.23**		**0.73**		**3.32**

According to the results in Table 7.5, Model 8 has the minimum MAPE value of 0.73% whereas Model 1 has the highest MAPE value of 31.64%. The results also indicate that every optimization strategy improves the forecasting accuracy of GM (1,1), grey-Verhulst and NGBM (1,1).

7.5 Conclusion

Forecasting which is the process of making predictions based on past and present data in order to cope with the uncertainty of future by providing information about probable events and their consequences is essential in decision making. The accuracy of forecasting depends on both the characteristics of data and the forecasting model used. The recent developments in the literature indicate that grey forecasting methods which are classified as modern methods of forecasting are quite successful especially for small sample time series. Health expenditure per capita and total expenditure in health are important indicators that reflect a country's health status and economic growth. Accurate estimation of the values of these indicators can guide efficient health care policy making and resource allocation.

The HCE data set used in this study consists of limited data. At this point, the most important feature of grey forecasting models is that it provides good prediction using limited data. This study presents eight different grey forecasting models based on three different strategies which are rolling mechanism, training data size optimization and parameter optimization to improve forecasting accuracy of grey prediction models.

Firstly, traditional GM (1,1) model and Grey-Verhulst model which is a special case of NGBM (1,1) are used for predicting HCE per capita. Then, in order to improve the forecasting accuracy of these two models, rolling mechanism, and training data size optimization and parameter optimization by using GA are applied. The comparison of these models based on MAPE measure indicate that the optimization of parameters and training data size together with rolling mechanism highly improve the forecasting performance. Training data size and parameter optimized NGBM (1,1) with rolling mechanism is the model with the highest forecasting accuracy. The results also indicate that the parameter optimization decreases the forecasting error more than the training data size optimization.

Optimized ARIMA model with rolling mechanism is also included in the study in order to compare the forecasting performance of the grey forecasting based models with statistical methods. Both of the optimized GM (1,1) and NGBM (1,1) with rolling mechanisms outperform the optimized ARIMA model with rolling mechanism.

These results show that the optimized NGBM (1,1) model with rolling mechanism gives robust results and it can be applied for other forecasting problems. The optimization of these presented grey forecasting models by using different meta-heuristic methods can be considered for further research.

References

Abdel-Aal RE, Mangoud AM (1998) Modeling and forecasting monthly patient volume at a primary health care clinic using univariate time-series analysis. Comput Methods Prog Biomed 56(3):235–247

Aboagye-Sarfo P, Mai Q, Sanfilippo FM, Preen DB, Stewart LM, Fatovich DM (2015) A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in Western Australia. J Biomed Inform 57:62–73

Akay D, Atak M (2007) Grey prediction with rolling mechanism for electricity demand forecasting of Turkey. Energy 32(9):1670–1675

Alvisi S, Franchini M (2012) Grey neural networks for river stage forecasting with uncertainty. Phys Chem Earth Parts A/B/C 42:108–118

Astolfi R, Lorenzoni L, Oderkirk J (2012) Informing policy makers about future health spending: a comparative analysis of forecasting methods in OECD countries. Health Policy 107(1):1–10

Bahrami S, Hooshmand RA, Parastegari M (2014) Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm. Energy 72:434–442

Benítez RBC, Paredes RBC, Lodewijks G, Nabais JL (2013) Damp trend grey model forecasting method for airline industry. Expert Syst Appl 40(12):4915–4921

Blanco-Moreno Á, Urbanos-Garrido RM, Thuissard-Vasallo IJ (2013) Public healthcare expenditure in Spain: measuring the impact of driving factors. Health Policy 111(1):34–42

Box GE, Jenkins GM (1976) Time series analysis: forecasting and control, revised edn. Holden-Day, San Francisco

Chang CJ, Li DC, Huang YH, Chen CC (2015) A novel gray forecasting model based on the box plot for small manufacturing data sets. Appl Math Comput 265:400–408

Chang SC, Lai HC, Yu HC (2005) A variable P value rolling grey forecasting model for Taiwan semiconductor industry production. Technol Forecast Soc Chang 72(5):623–640

Chatfield C (2000) Time-series forecasting. CRC Press, Florida, USA

Chen CI (2008) Application of the novel nonlinear grey Bernoulli model for forecasting unemployment rate. Chaos, Solitons Fractals 37(1):278–287

Chen CI, Chen HL, Chen SP (2008) Forecasting of foreign exchange rates of Taiwan's major trading partners by novel nonlinear grey Bernoulli model NGBM (1, 1). Commun Nonlinear Sci Numer Simul 13(6):1194–1204

Chen CI, Hsin PH, Wu CS (2010) Forecasting Taiwan's major stock indices by the Nash nonlinear grey Bernoulli model. Expert Syst Appl 37(12):7557–7562

Cui J, Liu SF, Zeng B, Xie NM (2013) A novel grey forecasting model and its optimization. Appl Math Model 37(6):4399–4406

Deng J (1982) Control problems of grey systems. Syst Control Lett 1(5):288–294

Deng J (1986) Grey prediction and decision. Huazhong University of Science and Technology, Wuhan

Diebold FX (1998) Elements of forecasting, 4th edn. South-Western College Publications, Ohio, USA

Eswaran C, Logeswaran R (2012) A dual hybrid forecasting model for support of decision making in healthcare management. Adv Eng Softw 53:23–32

Feng SJ, Ma YD, Song ZL, Ying J (2012) Forecasting the energy consumption of China by the grey prediction model. Energy Sources Part B Econ Plan Policy 7(4):376–389

Ferrand Y, Kelton CM, Guo JJ, Levy MS, Yu Y (2011) Using time-series intervention analysis to understand US Medicaid expenditures on antidepressant agents. Res Soc Adm Pharm 7(1): 64–80

Froelich W, Salmeron JL (2014) Evolutionary learning of fuzzy grey cognitive maps for the forecasting of multivariate, interval-valued time series. Int J Approx Reason 55(6):1319–1335

Getzen T (2000) Forecasting health expenditures: short, medium and long (long) term. J Health Care Finance 26(3):56–72

Gille L, Houy T (2014) The future of health care demand in developed countries: from the "right to treatment" to the "duty to stay healthy". Futures 61:23–32
Hamzacebi C, Es HA (2014) Forecasting the annual electricity consumption of Turkey using an optimized grey model. Energy 70:165–171
Hartwig J, Sturm JE (2014) Robust determinants of health care expenditure growth. Appl Econ 46(36):4455–4474
Holland JH (1975) Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. The University of Michigan Press, Ann Arbor
Hsu CC, Chen CY (2003) Applications of improved grey prediction model for power demand forecasting. Energy Convers Manag 44(14):2241–2249
Hsu LC (2009) Forecasting the output of integrated circuit industry using genetic algorithm based multivariable grey optimization models. Expert Syst Appl 36(4):7898–7903
Hsu LC (2010) A genetic algorithm based nonlinear grey Bernoulli model for output forecasting in integrated circuit industry. Expert Syst Appl 37(6):4318–4323
Hsu LC (2011) Using improved grey forecasting models to forecast the output of opto-electronics industry. Expert Syst Appl 38(11):13879–13885
Hsu LC, Wang CH (2007) Forecasting the output of integrated circuit industry using a grey model improved by the Bayesian analysis. Technol Forecast Soc Chang 74(6):843–853
Hsu YT, Liu MC, Yeh J, Hung HF (2009) Forecasting the turning time of stock market based on Markov–Fourier grey model. Expert Syst Appl 36(4):8597–8603
Huang KY, Jane CJ (2009) A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories. Expert Syst Appl 36(3):5387–5392
Huang Y, Wang Y, Gai S (2011) The application and research of a new combinatorial analysis and forecasting method in real estate area based on grey system theory and multivariate linear regression. Procedia Eng 15:4532–4537
Huarng KH, Yu THK (2014) A new quantile regression forecasting model. J Bus Res 67(5):779–784
Intharathirat R, Salam PA, Kumar S, Untong A (2015) Forecasting of municipal solid waste quantity in a developing country using multivariate grey models. Waste Manag 39:3–14
Jalalpour M, Gel Y, Levin S (2015) Forecasting demand for health services: development of a publicly available toolbox. Oper Res Health Care 5:1–9
Jiang X, Zhang L, Chen XM (2014) Short-term forecasting of high-speed rail demand: a hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China. Transp Res Part C Emerg Technol 44:110–127
Jiang Y, Yao Y, Deng S, Ma Z (2004) Applying grey forecasting to predicting the operating energy performance of air cooled water chillers. Int J Refrig 27(4):385–392
Kang J, Zhao H (2012) Application of improved grey model in long-term load forecasting of power engineering. Syst Eng Procedia 3:85–91
Kumar U, Jain VK (2010) Time series models (grey-Markov, grey model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. Energy 35(4): 1709–1716
Lago-Peñas S, Cantarero-Prieto D, Blázquez-Fernández C (2013) On the relationship between GDP and health care expenditure: a new look. Econ Model 32:124–129
Lee R, Miller T (2002) An approach to forecasting health expenditures, with application to the US Medicare system. Health Serv Res 37(5):1365–1386
Lee YS, Tong LI (2011) Forecasting energy consumption using a grey model improved by incorporating genetic programming. Energy Convers Manag 52(1):147–152
Lei M, Feng Z (2012) A proposed grey model for short-term electricity price forecasting in competitive power markets. Int J Electr Power Energy Syst 43(1):531–538
Li DC, Chang CJ, Chen CC, Chen WC (2012) Forecasting short-term electricity consumption using the adaptive grey-based approach – an Asian case. Omega 40(6):767–773

Li DC, Chang CJ, Chen WC, Chen CC (2011a) An extended grey forecasting model for omnidirectional forecasting considering data gap difference. Appl Math Model 35(10): 5051–5058
Li DC, Yeh CW, Chang CJ (2009) An improved grey-based approach for early manufacturing data forecasting. Comput Ind Eng 57(4):1161–1167
Li GD, Wang CH, Masuda S, Nagai M (2011b) A research on short term load forecasting problem applying improved grey dynamic model. Int J Electr Power Energy Syst 33(4):809–816
Lin YH, Lee PC, Chang TP (2009) Adaptive and high-precision grey forecasting model. Expert Syst Appl 36(6):9658–9662
Liu X (2013) A grey neural network and input-output combined forecasting model and its application in primary energy related CO 2 emissions estimation by sector in China. Energy Procedia 36:815–824
Ma W, Zhu X, Wang M (2013) Forecasting iron ore import and consumption of China using grey model optimized by particle swarm optimization algorithm. Res Policy 38(4):613–620
Maisonneuve C, Martins JO (2013) A projection method for public health and long-term care expenditures, OECD Economics Department Working Paper No. 1048. Organisation for Economic Co-operation and Development, Paris. http://dx.doi.org/10.1787/5k44v53w5w47-en
Marešová P, Mohelská H, Kuča K (2015) Economics aspects of ageing population. Procedia Econ Finance 23:534–538
Mun J (2006) Modeling risk: applying Monte Carlo simulation, real options analysis, forecasting, and optimization techniques, vol 347. Wiley, New York, USA
Munkin MK, Trivedi PK (2003) Bayesian analysis of a self-selection model with multiple outcomes using simulation-based estimation: an application to the demand for healthcare. J Econ 114(2):197–220
Ou SL (2012) Forecasting agricultural output with an improved grey forecasting model based on the genetic algorithm. Comput Electron Agric 85:33–39
Pao HT, Fu HC, Tseng CL (2012) Forecasting of CO 2 emissions, energy consumption and economic growth in China using an improved grey model. Energy 40(1):400–409
Prieto DC, Lago-Peñas S (2012) Decomposing the determinants of health care expenditure: the case of Spain. Eur J Health Econ 13(1):19–27
Ren XW, Tang YQ, Li J, Yang Q (2012) A prediction method using grey model for cumulative plastic deformation under cyclic loads. Nat Hazards 64(1):441–457
Samvedi A, Jain V (2013) A grey approach for forecasting in a supply chain during intermittent disruptions. Eng Appl Artif Intell 26(3):1044–1051
Sheng-qiang Y, Yan S, Zu-yun C, Bao-hai Y, Quan X (2009) Establishment of grey-neural network forecasting model of coal and gas outburst. Procedia Earth Planet Sci 1(1):148–153
Soyiri IN, Reidpath DD (2013) An overview of health forecasting. Environ Health Prev Med 18(1):1–9
Sun X, Sun W, Wang J, Zhang Y, Gao Y (2016) Using a grey–Markov model optimized by Cuckoo search algorithm to forecast the annual foreign tourist arrivals to China. Tour Manag 52: 369–379
Tang HWV, Yin MS (2012) Forecasting performance of grey prediction for education expenditure and school enrollment. Econ Educ Rev 31(4):452–462
Tsai SB (2016) Using grey models for forecasting China's growth trends in renewable energy consumption. Clean Techn Environ Policy 18(2):563–571
Tsaur RC (2010) The development of an interval grey regression model for limited time series forecasting. Expert Syst Appl 37(2):1200–1206
Tseng FM, Yu HC, Tzeng GH (2001) Applied hybrid grey model to forecast seasonal time series. Technol Forecast Soc Chang 67(2):291–302
Van Baal PH, Wong A (2012) Time to death and the forecasting of macro-level health care expenditures: some further considerations. J Health Econ 31(6):876–887
Wang CH, Hsu LC (2008) Using genetic algorithms grey theory to forecast high technology industrial output. Appl Math Comput 195(1):256–263

Wang X, Cai Y, Chen J, Dai C (2013) A grey-forecasting interval-parameter mixed-integer programming approach for integrated electric-environmental management–a case study of Beijing. Energy 63:334–344
Wang Z (2009) Stock returns and the short-run predictability of health expenditure: some empirical evidence. Int J Forecast 25(3):587–601
Wei G (2011) Grey relational analysis model for dynamic hybrid multiple attribute decision making. Knowl-Based Syst 24(5):672–679
Wu L, Liu S, Liu D, Fang Z, Xu H (2015) Modelling and forecasting CO 2 emissions in the BRICS (Brazil, Russia, India, China, and South Africa) countries using a novel multi-variable grey model. Energy 79:489–495
Wu L, Liu S, Yao L, Yan S, Liu D (2013) Grey system model with the fractional order accumulation. Commun Nonlinear Sci Numer Simul 18(7):1775–1785
Wu WY, Chen SP (2005) A prediction method using the grey model GMC (1, n) combined with the grey relational analysis: a case study on Internet access population forecast. Appl Math Comput 169(1):198–217
Xia M, Wong WK (2014) A seasonal discrete grey forecasting model for fashion retailing. Knowl-Based Syst 57:119–126
Xie F, Chen Z, Shang J, Fox GC (2014) Grey forecast model for accurate recommendation in presence of data sparsity and correlation. Knowl-Based Syst 69:179–190
Xie NM, Liu SF (2009) Discrete grey forecasting model and its optimization. Appl Math Model 33(2):1173–1186
Xie NM, Liu SF, Yang YJ, Yuan CQ (2013) On novel grey forecasting model based on non-homogeneous index sequence. Appl Math Model 37(7):5059–5068
Xie NM, Yuan CQ, Yang YJ (2015) Forecasting China's energy demand and self-sufficiency rate by grey forecasting model and Markov model. Int J Electr Power Energy Syst 66:1–8
Yao AW, Chi SC, Chen JH (2003) An improved grey-based approach for electricity demand forecasting. Electr Power Syst Res 67(3):217–224
Yu THK, Wang DHM, Wu KL (2015) Reexamining the red herring effect on healthcare expenditures. J Bus Res 68(4):783–787
Zeng XY, Shu L, Huang GM, Jiang J (2015) Triangular fuzzy series forecasting based on grey model and neural network. Appl Math Model. http://dx.doi.org/10.1016/j.apm.2015.08.009
Zhang H, Li Z, Chen Z (2003) Application of grey modeling method to fitting and forecasting wear trend of marine diesel engines. Tribol Int 36(10):753–756
Zhan-Li M, Jin-Hua S (2011) Application of grey-Markov model in forecasting fire accidents. Procedia Eng 11:314–318
Zhao Z, Wang J, Zhao J, Su Z (2012) Using a grey model optimized by differential evolution algorithm to forecast the per capita annual net income of rural households in China. Omega 40(5):525–532
Zhou J, Fang R, Li Y, Zhang Y, Peng B (2009) Parameter optimization of nonlinear grey Bernoulli model using particle swarm optimization. Appl Math Comput 207(2):292–299
Zhou PABW, Ang BW, Poh KL (2006) A trigonometric grey prediction approach to forecasting electricity demand. Energy 31(14):2839–2847

Chapter 8
The Impact of Social Networks in Developing and Managing Chronic Care Models

Marina Resta and Elisabetta Arato

8.1 Introduction

As a matter of fact, chronic disease is an international concern which causes worldwide, according to the WHO[1] report (World Health Organization and Alzheimer's Disease International 2012), about 60% of deaths, and absorbs about 75% of the healthcare resources of a country.

Nevertheless, it is a common practice, justified by budgeting costs, to cover hospital stays only for acute diseases, and for very short periods of hospitalization; the consequence is that home physicians must take charge of increasingly demanding patients, as their problems are generally physical, psychological and social, altogether interacting. The key role of primary care is therefore to integrate the contributions of different professionals, thus offering a coherent and sustainable assistance, customized according to the context and capabilities of the individual patient. The Chronic Care Model –CCM (Wagner et al. 1996; Wagner 1998) is an example of this type of approach: it assumes a multi-pronged strategy covering six interrelated components[2] that should turn into a network where informed and activated patients interact with prepared and proactive practice-teams. Pilot case studies applying the CCM to the care of people with diabetes (McCulloch et al. 1998; Ramsey et al. 1999) revealed the potential of the approach. The model has

[1]World Health Organization.

[2]Self-management support, clinical information systems, delivery system redesign, decision support, health care organization, and community resources.

M. Resta (✉) • E. Arato
Dipartimento di Economia, Università degli Studi di Genova, Genoa, Italy
e-mail: resta@economia.unige.it; earato@economia.unige.it

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_8

been also implemented by a large number of organizations in the United States, United Kingdom and Sweden with encouraging results (Wagner et al. 2001).

Following this successful trail, the Italian law 189/2012[3] promoted a new approach to manage primary care, with a model of chronic care, whose main aim is to ensure the patient with proper actions, differentiated according to the level of risk, also focusing on prevention and education. In the Italian experience, the main issues of the model concern the poor level of integration often affecting the community of family physicians, hospitals and other providers in terms of culture, organization and financing: while these independent arrangements may offer some benefit, such as broadened patient choice, the flip side is fragmentation across care sites, providers and in clinical decision making for patients.

We first focus on a population benchmark of patients affected by chronic degenerative disease, then we provide evidence of how SNA can help to understand the dynamics of mutual relationships between the network of home -physicians and the other entities within the National Health System, and to suggest corrective interventions to improve the overall quality of the provided service.

The work is organized as follows: Sect. 8.2 illustrates the literature debate, while Sect. 8.3 depicts the framework for our study, giving details on the employed data; Sect. 8.4 discusses the results. Finally, Sect. 8.5 concludes.

8.2 The Literature Debate

8.2.1 The General Framework

Over the past 20 years a research wave focused on the development of inter-organizational networks within the government sector. Expressions such as "hollow State", "third party government", and "market state", have become of common use to characterize the observed increase in productivity, and in the dependence of the public to the private markets, as well as the broad decentralization process, in search for new tools to assess the accountability of results. Within this process a decisive role has been played by the network paradigm. The research interest for this topic has remote origins and belongs to different disciplines: in the early 1990s of the twentieth century (Alter and Hage 1993) has defined the network as:

> a set of actors who are interdependent in order to achieve a goal, that gives rise to a collective output (tangible – intangible), that no actor could individually get.

Here the term actor can be easily replaced with organization. Furthermore, (Berry et al. 2004) referred to process of *"cross-fertilization"*, which includes the fields of sociology, politics and public management, while (Weber and Khademian 2008)

[3] http://www.altalex.com/documents/leggi/2012/11/12/dl-sanita-la-legge-di-conversione-del-decreto-balduzzi

focused on the study of ongoing relationships involving organizations, individuals and groups, in order to solve complex problems, cope with the scarcity of resources and achieve goals. A similar vision is shared by (Milward et al. 2009), looking at networks as alternatives when markets fail. Another definition of network may be found in (Porter and Van der Linde 1995) as the strategy to create value and competitive advantage, while (Barringer and Harrison 2000) used the metaphor of the constellation of organizations to interpret the definition of network. In (Rhodes 1996) the network is portrayed as a form of *"mediation of interests"*; aligned to this vision, (Klijn and Koppenjan 2000) and (Klijn et al. 2010) focused on the overall coordination of the relationships that have been created. Finally, part of the existing literature interprets the network as a possible form of governance: this is the case for instance of (Kickert and Koppenjan 1997):

> the network gathers more or less stable social relationships between interdependent actors, which take shape around political issues or political agendas.

For what is concerning the Italian literature, (Butera 1990) provides a discussion of the elements building the network: the nodes are viewed as result-oriented entities, relatively self-regulated, and able to cooperate with others. They have their own *vitality*, i.e. the ability to survive independently and to communicate with other systems for exchange of information and values), the connections (the ties, the relationships between the nodes) and finally the structure (the set of nodes and connections).

In all the discussed cases, we can identify some common elements that well specify the network: social interaction, relationships, connections, partnerships, collective action, confidence, cooperation. Nevertheless, the theme appears to be very broad; in order to work within a narrower framework, we refer to the classification proposed by (Isett et al. 2011), hence focusing on collaborative networks.

8.2.2 The Primary Care Stream

The expression "primary care" was used for the first time in the UK within the s.c. "Dawson report" (Dawson 1920) identifying the three levels of health care (primary, secondary, tertiary) that are still present in most national health systems. Later in the past century, the Alma Ata Conference in 1978 was the first international conference on primary healthcare, gathering worldwide expert to promote and develop comprehensive health-care services for all citizens, aimed to guarantee: (*i*) a form of equity in the implemented health strategies; (*ii*) a preventive approach; (*iii*) continuity in the activities (care or assistance); (*iv*) the joint involvement of all the figures around the patient. Furthermore, starting from early 2000s, a set of guidelines emerged towards which reforming the primary care.

A recent study by the World Health Organization (World Health Organization and Alzheimer's Disease International 2012) has shown that health systems focused

on primary care, have better outcomes, greater equity and continuity in performance, compared to the related systems on specialist care. In Italy the D. Lg. 158/2012 converted into Law 189/2012, better known as "law Balduzzi" highlighted the need for a healthcare system relying on primary healthcare. In particular the Law promotes a model for chronic care, whose main aim is to ensure the patient with proper actions, differentiated according to the level of risk, also focusing on prevention and education. This is a new model of care for a proactive taking charge of citizens and a new organizational approach, which assumes the need for health before the disease arises, hence creating a system around citizen to maintain the highest possible level of health, slowing the progression of chronic disease and preventing the onset of acute illness. In practice, the new system anticipates either the arising of the diseases or their progression, thus ensuring the patient appropriate interventions, differentiated according to the level of risk.

This model is based on a multidisciplinary approach, which places the patient at the center of the system, integrating the different actions related to prevention, diagnosis and treatment to rehabilitation-assistance long-term, all with a view to continuity of care. The health initiative is therefore based on the interaction between the patient and a team of experts covering different aspects of the health-care process.

8.2.3 The Chronic Care Model and Beyond

According to (World Health Organization and Alzheimer's Disease International 2012), chronic diseases, worldwide, are responsible for about 60% of deaths and account for about 75% of the healthcare resources of a country. The situation for Italy is summarized in Table 8.1.

Table 8.1 Chronic diseases as % of total population in 2013

Group ID	Age	Males	Females	Overall
1	0–14	8.9	6.5	7.7
2	15–17	14.1	13.1	13.6
3	18–19	14.5	16.6	15.5
4	20–24	14	19.3	16.7
5	25–34	16	15.3	15.7
6	35–44	22.3	22.3	22.3
7	45–54	35.9	39.1	37.5
8	55–59	53.8	56.2	55
9	60–64	61	66.4	63.7
10	65–74	72.8	78.1	75.6
11	>75	82.8	88.6	86.4

Source: Istituto Superiore di Statistica (ISTAT) 2015 database

The table shows the percentage impact of chronic diseases organized by the age groups given in Column 2. For each group we provided the score for the population group as a whole (Column 5) as well as for both males (Column 3) and females (Column 4). In order to make easier the discussion, Column 1 labels each age group by a positive integer number, where 1 identifies the first age group (0–14), and 11 the group gathering people older than 75. Looking at the scores, chronic diseases becomes significant (involving over 50% of the whole population) in the age group labelled by 8, with peaks for population of over 75. Moreover, it is expected a grow-up in this records for the future, due to increased life expectancy of people, which is strongly correlated to both the emergence of chronic diseases, and to the increase of health spending. Furthermore, according to the ISTAT database, the top ten causes of death in Italy are closely related to the emergence of chronic diseases: some examples are provided by the ischemic diseases, cerebrovascular diseases, often arising from hypertension and Alzheimer's and other dementia diseases. Faced with this scenario, the management of chronic cases become a major challenge for the National Health Systems that has variously taken inspiration from existing chronic care models -CCM, with a focus on the model of (Wagner et al. 1996).

The Wagner's model assumes the need for health before the disease arises, and hence creates a system around citizens allowing them to maintain the highest possible level of health, slowing the progression of chronic disease and preventing the onset of acute illness. This model is based on a multidisciplinary approach aimed at the chronic disease management, which places the patient at the center of the system, and tries to integrate different interventions, related to prevention, diagnosis and treatment to rehabilitation-assistance long-term, all with a view to continuity of care. The model relies on six key elements: the exploitation of the resources of the community, the health care organization, the support for self-care, the organization of the team (pro activity interventions), information and decision support systems (see Fig. 8.1 for more details).

The CCM gathers two different categories: community and health, with the first incorporating the second. The interaction of the elements building up the CCM, and namely: between the informed patient and the proactive team, makes possible to achieve better results.

An advanced version of the CCM is the Expanded Chronic Care Model -ECCM- due to (Barr et al. 2003), explained in Fig. 8.2: the patient is merged into the broader dimension of the community with clinical aspects gathered to more general considerations of public health, such as primary prevention and collective attention to the determinants of health. In this way, the ECCM involves not only citizens through promotion of health, health education and support to self-management, but also introduces a higher dimension of social responsibility, with citizens acting in the care pathway and results giving benefits to both the single patient and the whole society.

In both the CCM and ECCM the central role is played by the generic physicians, who, in most cases must be organized into modules combining generic physicians, specialists, nurses, who work in a multidisciplinary team using specially crafted care

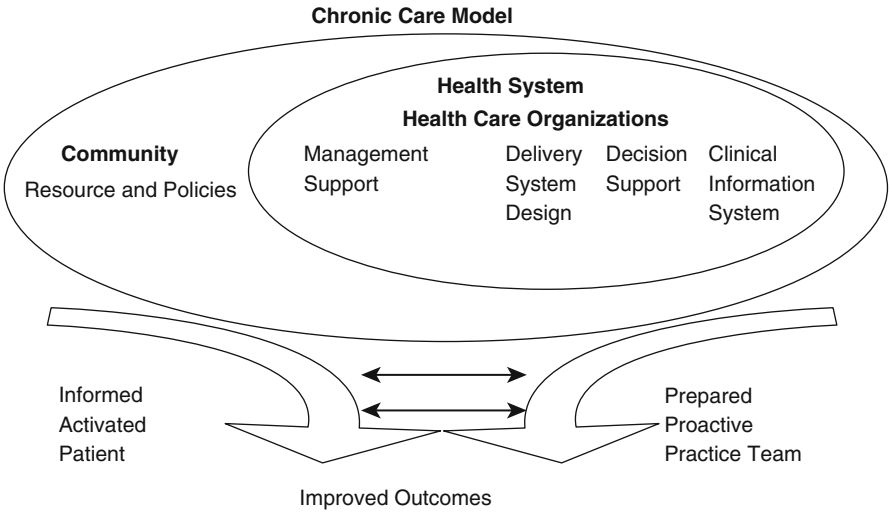

Fig. 8.1 The chronic care model

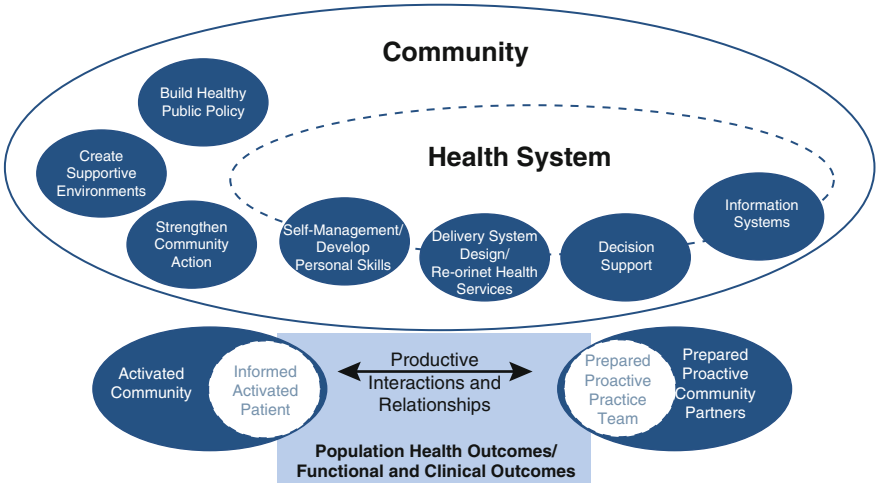

Fig. 8.2 The expanded chronic care model. Source: http://www.longwoods.com/content/16763

pathways and clinical records. In addition to health care professionals and to the patient, the process involves also local health authorities such as the municipality, the service companies, the District and of course the patient.

In general, those approaches testify the tendency of the health sector to customize and humanize health care services: the reduction of delays and the optimization of available resources, as well as the improvement of reached outcomes are the major results.

8.3 Case Study

According to (World Health Organization and Alzheimer's Disease International 2012) the pathology of dementia has a severe priority among the chronic diseases. The world data on dementia, aligned to the sharp increase in the rate of aging of the population, are dramatic: over 47,500,000 people with dementia, and 7.7 million new cases each year (in practice: one new case every 4 s); moreover, it is expected that in 2030 the global statistics will reach 75.6 million cases, to triple within 2050. The estimated cost for this condition is about $604 billion, about 1% of world GDP. In Italy patients with dementia are about 1.3 million over a population of 56,000,000 inhabitants: depending on the severity of the disease, the estimated costs per patient ranging between 9,000 and 16,000 Euro, for a total of 12.10 billion Euro.

The WHO has established a set of guidelines to be adopted in order to better manage this global emergency that can be summarized in the following points: (*i*) considering dementia a health priority for all countries; (*ii*) improving the attitude and knowledge of both common people and professionals with respect to dementia; (*iii*) investing in health and social systems to improve care and services for patients and their caregivers; (*iv*) increasing the priority given to dementia research agenda of public health. In the light of this dramatic scenario, and aligned to the path traced by the WHO, the long-awaited National Plan Dementias (Di Fiandra et al. 2015) has been released in Italy in 2014.

The need to create health-care networks stems from the necessity of properly responding to the increase in complexity of health needs, creating synergies in clinical processes, sharing clinical best practices, increasing levels of quality of provided services, containing costs, efficiently and effectively managing the impact of new technology in health-care. In order to do this, we focused on the Primary Care investigating the role of general practitioners, as a kind of *gatekeepers*, connecting the world of health care to both the territory and the hospital context. The general practitioner is the one who meets the patient at first, personally knows his clinical history, directs him (if necessary) to hospital or local health contexts. The function of the general practitioner is thus to provide a holistic view of the patient, promoting continuity of provided care and enabling an integration among various provided services.

Our research work is therefore inspired to an approach combining business management and quantitative methods. For the former, we have been mainly driven by the theoretical framework as described in (Cicchetti and Mascia 2007) and (Cicchetti et al. 2006), highlighting the set of steps making possible to apply Social Network Analysis -SNA- (Scott 1991; Wasserman and Faust 1994) within the managerial context. For the latter, we mainly refer to the framework depicted by (Tumminello et al. 2005) for financial markets and already applied to the health-care for analyzing related costs by (Resta 2012, 2013, and 2016).

8.3.1 Data and Methods

Collected data arose as the results of an online questionnaire that has been submitted to 98 generic physicians in four different Italian regions: Tuscany, Lombardy, Piedmont and Liguria. The sample represents a cross-section of Northern and Central parts of Italy: while earlier three regions are commonly recognized as excellence in managerial health, the latest one (Liguria) has been considered as it is known for being the region in Italy with the highest level of seniors: the ISTAT data for 2015 highlighted that 28% of the population in Liguria was over 65 (the Italian average is 21%, while the corresponding data for Europe is 18.5%.

The questionnaire was organized into five parts, with 99 questions including:

- Demographic aspects, with questions aimed to determine the age of the physician, the university of graduation, the specialization (if owned), as well as the number of years since he has been working as generic physician.
- Issues related to the management of chronic patients. Here the questions were oriented to quantify the proportion of chronic patients, with a focus on the number of visits, the degree of compliance of the patient with respect to the prescriptions given by the physician, the use of indicators to monitor patients' chronic level.
- Relational aspects: the physician were asked to identify the main actors in their professional network, the factors facilitating the creation of the professional network, as well as to evaluate the level of integration of the different network agents, and the degree of balance with respect to the assigned tasks in managing chronic patients.
- Sources of information for practitioners: the physician were asked to indicate the main sources obtaining information related to chronic diseases (Dandi et al. 2013).
- Issues on proactive approaches in health.

With the collected data (a matrix 98×99), we assumed to manage each respondent as a network node identified by an alpha-numerical string (to preserve his anonymity). The matrix dimensions are a consequence of the questionnaire structure: each row represents a physician (98 elements in the sample) and each column a score addressing the answers to the 99 provided questions.

We then evaluated the correlation ρ for each couple of nodes (as to say: each couple of physicians), building a 98×98 adjacency matrix $A^* = \{a_{ij}\}$, with $a_{ij} = 1$, if $\rho_{ij} \neq 0$, and $a_{ij} = 0$, otherwise. The graph representing the correlation structure among nodes is presented in Fig. 8.3.

The graph in Fig. 8.3 highlights the large number of connections among the nodes, and hence the difficulty in interpreting the excess of provided information. For this reason, we preferred to prune the graph, extracting the essential features of the network. We applied two different approaches, respectively based on: (*a*) correlation filtered matrices; (*b*) a graph filtering algorithm, leading to Planar Maximally Filtered Graphs (PMFG).

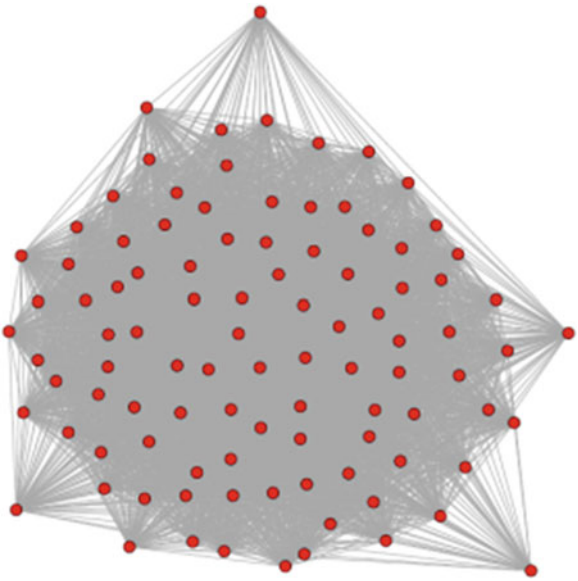

Fig. 8.3 The correlation structure of the physicians' network

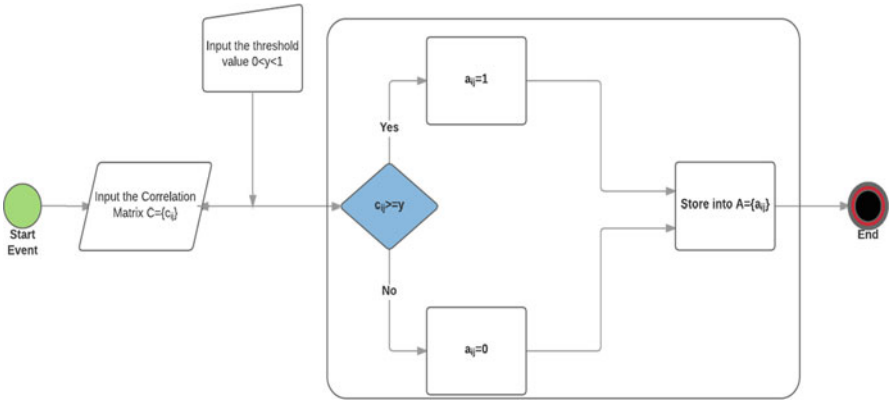

Fig. 8.4 Correlation filtered networks: the basic procedure

The two procedures are illustrated in Figs. 8.4 and 8.5, where we provided the corresponding algorithmic schemes that lead to networks filtered in accordance to either the correlation (Fig. 8.4) or the PMFG (Fig. 8.5).

Understanding how a correlation filter works is quite trivial: once a level of correlation (say for instance: $y = 0.5$) has been set, then all the ties among nodes with correlation below this threshold value are discarded. On the other hand, the PMFG (Tumminello et al. 2005) is worth of a few more words spending. In this case, the correlation values are initially ordered in decreasing order and gathered into a list S. Then, the first tie of the network is established, by connecting the

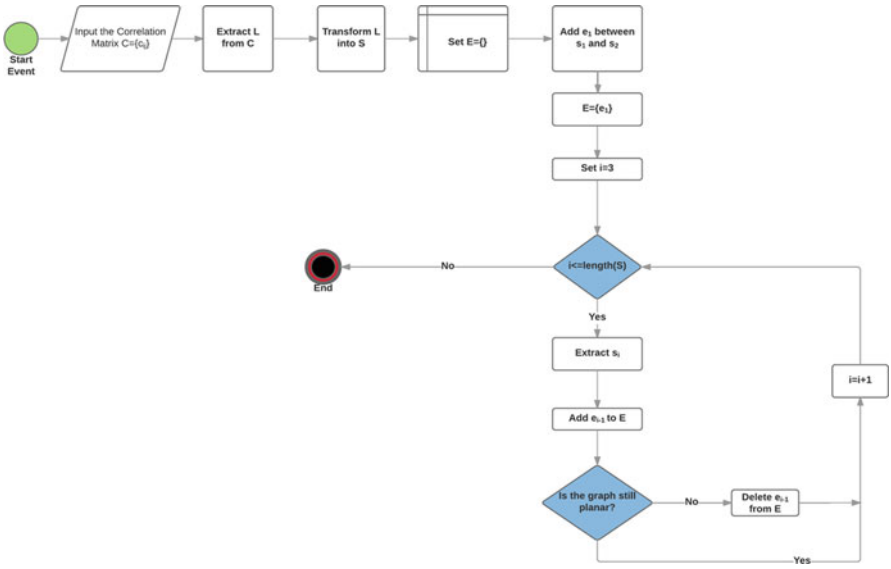

Fig. 8.5 Algorithmic scheme for the planar maximally filtered network

nodes with the highest correlation (and corresponding to earlier two elements in S), while any further connection (following the ordering in S) is added if and only if it preserves the planarity features of the emergent network. The PMFG has been then properly described as a method to simplify the network of interaction, by keeping only the most significant interactions (largest correlation) but in such a way that simultaneously extracts global information about the hierarchical organization of the whole system.

8.4 Discussion of Results

As said in previous section we started to analyze data by way of correlation filtered networks, varying the threshold level in the range from 0.2 to 0.7: Figure 8.6 shows the resulting networks.

The rationale of introducing a threshold level y may be searched in the need of moving from a very complex set of ties (likewise in Fig. 8.6a) towards a set of more significant and readable linkages (likewise in Fig. 8.6b and *a fortiori* in Fig. 8.6c, d). As a matter of fact, as the level of correlation into the observed sample is high (over 0.4) for all the nodes, keeping the threshold value below 0.5 did not guarantee to skim essential features of data. For the same reason, we avoided to represent networks at the correlation threshold of both 0.3 and 0.4, because the resulting linkages were the same as in the case of $y = 0.2$ (Fig. 8.6a). Moreover, as the declared aim of this task was retrieving more relevant information from the data,

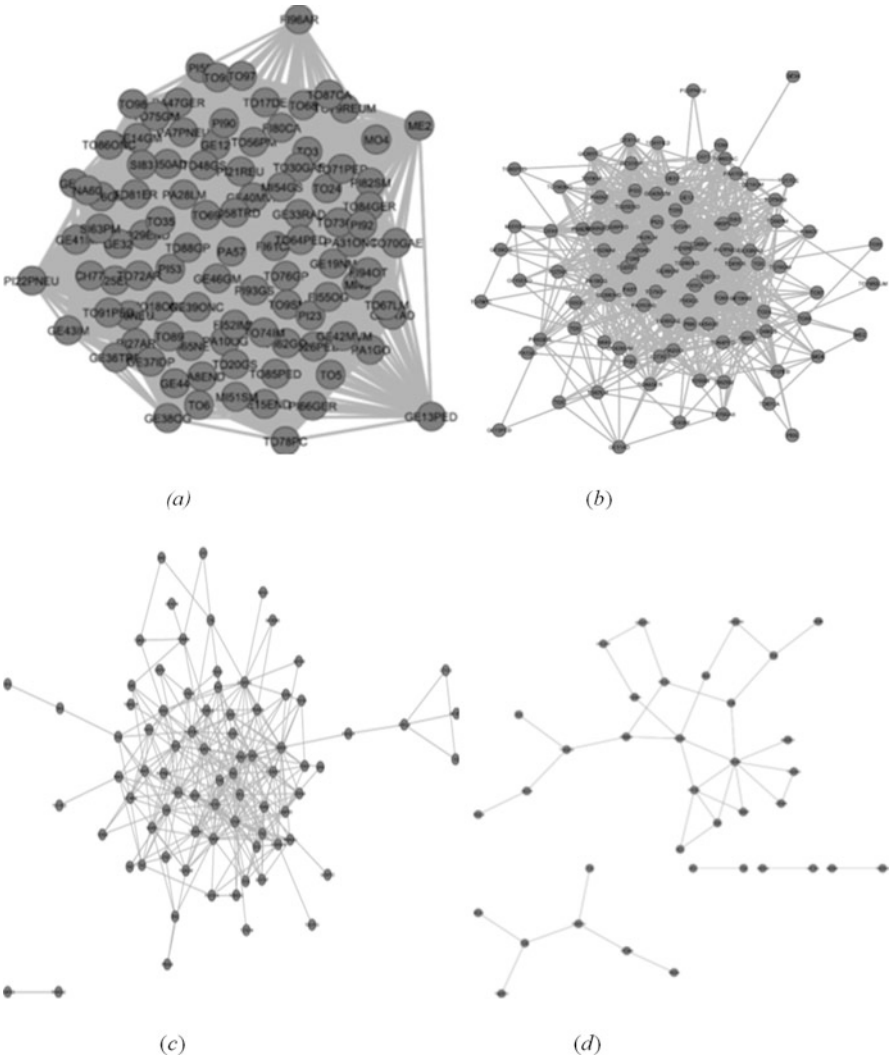

Fig. 8.6 Correlation filtered networks with threshold y = 0.2 (**a**); y = 0.5 (**b**); y = 0.6 (**c**), and y = 0.7 (**d**)

the goal was missing at this correlation level as well as for every y < 0.5. Moving to the network obtained by setting the threshold level at least equal to 0.5 (Fig. 8.6b), the structure begins to highlight more relevant ties, and the same happens if we turn to Fig. 8.6c, d representing the network structure at y = 0.6 and y = 0.7, respectively.

Unfortunately, as the threshold level increases together with the emergence of very relevant structure, by construction, the number of deleted nodes increases, so that the network loses its representation and generalization capabilities.

Table 8.2 Trade-off between the value of the correlation filter and the number of connected nodes in the network

Correlation threshold	Deleted nodes	Connected nodes
0.2	0	98
0.3	0	98
0.4	0	98
0.5	2	96
0.6	18	80
0.7	53	45
>0.7	98	0

Fig. 8.7 The PMFG representation of the physicians' network

Table 8.2 reports the summary of the number of nodes which are eliminated as the threshold level of the correlation increases: while up to the correlation level $y = 0.40$ all nodes are maintained, on the other hand, when the threshold y exceeds 0.7 all nodes become disjoint.

In order to avoid this loss of representativeness, the natural step was to consider a filtering alternative, letting unchanged the number of nodes. The choice of PMFG has been already motivated in previous Sect. 8.3, so that we are now focusing on the results discussion only. Figure 8.7 shows the network obtained by applying the PMFG algorithm to our data.

Moreover, we calculated and analyzed some Social Network Analysis (SNA) indexes on all the graphs arising from both correlation-filtered matrices and the PMFG. The choice of the indicators to use was not so immediate, as they are very numerous. It was decided to consider some measures of connection and centrality that are described in next rows.

Table 8.3 Some SNA indexes for the examined networks

Network ID	DEN	CC	ACV
Corr 20	0.91269	0.93358	0.92408
Corr 30	0.75657	0.84100	0.81104
Corr 40	0.50999	0.70768	0.65349
Corr 50	0.25197	0.56602	0.42799
Corr 60	0.10222	0.39487	0.27981
Corr 70	0.06190	0.18072	0.17756
PMFG	0.06059	0.36287	0.25163

- Density (DEN), measured as the ratio between the number of links between the nodes and the maximum number of possible relationships. In practice, DEN is a measure of connection: the closer to one, the more the number of edges (links) is approaching the maximal number of linkages.
- Clustering Coefficient (CC), which measures how much the nodes of a network might be involved into a cluster (a group). Also in this case, when CC values go closer to one, they generally indicate a strong tendency to clusters formation, and vice versa as values approach zero.
- Average Centrality Value (ACV), which indicates the average value of the bonds that each node can have on the network, and hence can be used to measure the strength of the neighborhood for each node and to eventually support the evidence provided by the CC.

Table 8.3 shows the value obtained.

The density exhibits a downward trend for networks filtered by correlation, as the threshold value increases: this is not surprising, since it is a consequence of eliminating knots: in the network with threshold at 0.20 the density has a very high value (about 0.91), as the nodes are richly interconnected, while in PMFG, the density assumes a low value (very close to the one obtained filtering at the correlation level of 0.70). Interestingly, while this result in correlation filtered networks is obtained by eliminating nodes, on the other hand in PMFG the same goal is reached by eliminating not so relevant information. Moving to the clustering coefficient, it assumes decreasing value in correlation filtered networks, while in the case of PMFG it has an intermediate value between those obtained by Corr 50 and Corr 60 networks. For what is concerning the average value of centrality, similar considerations to those already made in the case of the clustering coefficient apply.

A preliminary conclusion is then that the PMFG confirms to offer a superior way in representing data, as it maintains all the nodes but eliminating redundant connections, thus giving importance to the very features of examined data.

Starting from this point, we then grouped the PMFG nodes to let the natural clusters (Girvan and Newman 2002) to emerge, obtaining the physicians organization which is shown in Table 8.4: each physician is identified by a string whose earlier two elements identify the region where he graduated, while the remaining letters (when present) identify his major specialization; finally, the numbers register his physical order into the database. To make an example, PAPNEU7 identifies a physician who graduated in the town of Palermo, holding a specialization in

Table 8.4 Clusters composition in the PMFG network

CL01	CL02	CL03	CL04	CL05
TO6	GE12	ME2	TODE17	PAGO1
PIPNEU22	GEGM14	TO3	GE32	TO5
PAONC31	PIREU21	MO4	GE34	PAEND8
MI45	PI23	PAPNEU7	TO35	PAOG10
TOPM56	TOEND29	TOSM9	GETRD36	GEAD11
NA60	TOGAE30	TOGM16	GEIDP37	GEPED13
CH77	GERAD33	GENM19	GEMVM40	GEEND15
SI83	GEGM46	TO24	GEIM41	TOONC18
PIGS93	MINEU49	GEIM46	GE44	TOGS20
	MIGS54	TOGS48	PAGER47	TOEND25
	FIOG55	TOTRD58	PI53	TOPED26
	TO68	PI59	PINE65	PIAR27
	TOAR72	FICA61	TOONC86	PALM28
	TOGS73	FIGO61	TOPED91	GEOG38
	TOGM75	SIPM63		GEONC39
	TOER81	TOPED64		GEMVM42
	PISM82	TO69		FIAD50
	TOGER84	TOGAE70		MISM51
	TOGP88	TOPED71		FIIM52
	TO89	TOGP76		PA57
	PI92	TOREUM79		PIIGER66
	TO95	FICA80		TOLM67
	FIAR96	TOCA87		TOIM74
	TO98	PI90		TOPC78
		TO97		TOPED85
				FIOT94
9	24	25	14	26

Pneumology, and materially recorded as the 7-th entry in our database (in a word: he was the seventh physician registering to answer to our questionnaire). The complete list of specializations shortcuts is given in Appendix 1.

Looking to Table 8.4, we can observe that the first cluster consists of nine nodes, with no doctors from Liguria. This cluster is characterized by a low average age (50.78 years), but a relatively higher standard deviation (13.13). The values referring to both the years after graduation, and the years since starting the practice of the profession are lower than the average of the other clusters. Over 67% of physicians in this cluster do not belong to a medicine group: spatial proximity, specific features of the person and confidence, appear to be the elements which seem to facilitate the creation of a network; furthermore, within their professional network, the specialist is identified as the most important reference figure. The physicians in this cluster have the highest rate of use of specific indicators to monitor their chronic patients; nevertheless, only a reduced number of their chronic patients seem to

respect their medical prescription. Moving to Cluster 2, it is formed by 24 doctors, homogeneously belonging to the four regions of the observed sample. This cluster gathers physicians belonging to medical groups, with high percentage of visits to chronic patients. These physicians do not consider key factors for creating a network neither similarity of age or specialization. Cluster 3 is made up by 25 physicians, mostly from Piedmont. They are characterized by a low use of specific indices to monitor chronic patients. A specific feature of this group is that all physicians in it adhere to the dementia protocol for patients. The fourth cluster consists of 14 doctors, mostly from Liguria. Also in this case, all physicians belong to a medicine group. Physicians in this group use other doctors as source of information, while little value is given to both official documents and pharmaceutical information. Furthermore 93% of the elements in the cluster (the highest percentage on the five groups), knows about the existence of protocols for chronic patients. Coherently, there is a high satisfaction rate for the way chronic patients are managed.

Finally, Cluster 5 groups 26 physicians whose age on average has highest value (60.73 years). This group is characterized by a low number of chronic patients taken into charge, as well as by physicians giving higher value to the role of nurse, health professionals and local health authorities.

Looking at these results, PMFG seemed able to provide insights concerning the effective penetration of the prescription of the law into the network of generic physicians: our results, in fact, show that the formation of groups of physicians more permeable to the application of care pathways depends on the age of the individuals as well as by their geographical distribution. Moreover, in general physicians more inclined to apply new protocols are also those more up-to-date, giving great importance to the central role of chronic patients.

8.5 Conclusion

The main objective of this work was trying to offer a reliable and representative synthesis of multiple healthcare data through Social Network Analysis. SNA was used to investigate the role of the generic physician, within his professional network for the management of chronic patients. To this aim we started by highlighting the main issues faced by national health systems including population aging, the reduction of available resources, and the increasing incidence of chronic diseases, hence evidencing the great importance of the patient's centrality, in order to define new satisfactory health policies.

Our contribution mainly resides on two aspects.

The novelty of our work primarily stands on the developed questionnaire, aimed to investing in all their facets the aspects either promoting or discouraging the emergence of working groups in managing chronic patients.

The second element consists in the use of SNA methods to evaluate the answers to the questionnaire. Although there is a growing number of technical works already dealing with the application of network paradigms to the health care (Scott et al.

2005; Mascia 2009; Di Vincenzo et al. 2014; Munoz et al. 2014; Swan 2009; Tasselli 2014), the use of both correlation-filtered networks and Planar Maximally Filtered Graphs (PMFG) is a kind of first time application.

The results we obtained concern a sample involving 98 generic physicians, taken as representative for a cross-section of Northern and Central parts of Italy, including both regions that are commonly recognized as excellence in managerial health, and a region with the highest level of seniors in Italy. We show that, despite from the imposition by law, there is still a great number of generic physicians who continue not to use, or even not to know about what the new National Dementia Plan assumes: it is still a long way from theory to daily practice. Clearly by enlarging the observed sample it could be possible to achieve a better representation of the current situation.

Finally, our attempt may be considered as the first step to develop a quantitative protocol of more generic use, as its application might be broadened to specialists, to study and map out their function within the care path.

A.1 Appendix 1

List of shortcuts employed to abbreviate specializations names.

Name	Shortcut
Gynecology and obstetrics	GO
Obstetrics	OG
Pneumology	PNEU
Endocrinology	END
Sports medicine	SM
Digestive system apparatus diseases	AD
Pediatrics	PED
Generic medicine	GM
Dermatology	DE
Oncology	ONC
Nuclear medicine	NM
Generic specialization	GS
Rheumatology	REU
Anesthesia and intensive care	AR
Occupational medicine	LM
Gastroenterology	GAE
Respiratory apparatus diseases	RAD
Tisis and other respiratory apparatus diseases	TRD
Infectious diseases and pediatrics	IDP
Medical microbiology and virology	MVM
Internal medicine	IM

(continued)

Name	Shortcut
Geriatrics	GER
Emergency surgery and first aid	GS
Neuropathology	NEU
Immunology	IMM
Prevention medicine	PM
Cardiology	CA
Nephrology	NE
Generic surgery	GS
Generic pathology	GP
Palliative care	PC
Emergency room and first aid	ER
Otolaryngology	OT
Anesthesia and intensive care	AR

References

Alter C, Hage J (1993) Organizations working together. Sage, Newbury Park

Barr V, Robinson S, Marin-Link B, Underhill L, Dotts A, Ravensdale D, Salivaras S (2003) The expanded chronic care model. Hosp Q 7(1):73–82

Barringer BR, Harrison JS (2000) Walking a tightrope: creating value through interorganizational relationships. J Manag 26(3):367–403

Berry FS, Brower RS, Choi SO, Goa WX, Jang H, Kwon M, Word J (2004) Three traditions of network research: what the public management research agenda can learn from other research communities. Public Adm Rev 64(5):539–552

Butera F (1990) Il castello e la rete. Franco Angeli Libri, Milano

Cicchetti A, Mascia D (2007) Organizzare le reti in Sanità: teoria, metodi e strumenti di social network analysis. Mecosan 61:9–32

Cicchetti A, Papini P, Ruggeri M, De Luca A, Mascia D, Cippoloni E, Guasticchi G (2006) L'analisi dei network organizzativi nei sistemi sanitari: il caso della rete di emergenza nella Regione Lazio. Politiche Sanitarie 7(2):77–90

Dandi R, D'Alise C, Mascia D (2013) Network professionali e uso delle lenee guida: il caso della pediatria di libera scelta. Mecosan 87:9–22

Dawson B (1920) Interim report on the future provision of medical and allied services 1920. Ministry of Health, London

Di Fiandra T, Canevelli M, Di Pucchio A, Vanacore N, Italian Dementia National Plan Working Group (2015) The Italian dementia plan. Early release article. Ann Ist Super Sanità 51(4):261–264

Di Vincenzo F, Mascia D, Numerato D, Salvatore D (2014) Le reti sociali e i professionisti in sanità: una review della letteratura. Mecosan 89:71–83

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci U S A 99(12):7821–7826

Isett KR, Mergel IA, LeRoux K, Mischen PA, Rethemeyer RK (2011) Networks in public administration scholarship: understanding where we are and where we need to go. J Public Adm Res Theory 21(1):157–173

Kickert WJ, Koppenjan JF (1997) Public management and network management: an overview. Netherlands Institute of Government, Enschede
Klijn EH, Koppenjan JF (2000) Public management and policy networks: foundations of a network approach to governance. Public Manag Int J Res Theory 2(2):135–158
Klijn EH, Steijn B, Edelenbos J (2010) The impact of network management on outcomes in governance networks. Public Adm 88(4):1065
Mascia D (2009) L'organizzazione delle reti in sanità. Teoria, metodi e strumenti di social network analysis. FrancoAngeli, Milano
McCulloch DK, Price MJ, Hindmarsh M, Wagner EH (1998) A population-based approach to diabetes management in a primary care setting: early results and lessons learned. Eff Clin Pract. 1998 1(1):12–22
Milward HB, Provan KG, Fish A, Isett KR, Huang K (2009) Governance and collaboration: an evolutionary study of two mental health networks. J Public Adm Res Theory 20(1):125–141
Munoz DA, Alonso W, Nembhard HB (2014) A social network analysis-based approach to evaluate workflow and quality in a pediatric intensive care unit. In: Proceedings of the 2014 Industrial and Systems Engineering Research Conference
Porter ME, Van der Linde C (1995) Toward a new conception of the environment-competitiveness relationship. J Econ Perspect 9:97–118
Ramsey SD, Newton K, Blough D, McCulloch DK, Sandhu N, Wagner EH (1999) Patient-level estimates of the cost of complications in diabetes in a managed-care population. PharmacoEconomics 16(3):285–295
Resta M (2012) An analysis of the financial statements of Italian health care providers through correlation-based networks. In: Tànfani E, Testi A (eds) Advanced decision making methods applied to health care, vol 173. International Series in Operations Research & Management Science, Springer, Milan, pp 155–171
Resta M (2013) Social networks design of hospital facilitators: an empirical study of the case of Italy. J Mod Acc Audit 9.2. ISSN:1548-6583:1–21
Resta M (2016) Computational intelligence paradigms in economic and financial decision making. Springer, Switzerland
Rhodes RAW (1996) The new governance: governing without government. Political Stud 44(4):652–667
Scott J, Tallia A, Crosson JC, Orzano AJ, Stroebel C, Dicicco-Bloom et al (2005) Social network analysis as an analytic tool for interaction patterns in primary care practices. Annals Fam Med 3(5):443–448
Scott J (1991) Social network analysis. A handbook. Sage, London
Swan M (2009) Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 6(2):492–525
Tasselli S (2014) Social networks of professionals in health care organizations a review. Med Care Res Rev 71(6):619–660
Tumminello M, Aste T, Di Matteo T, Mantegna RN (2005) A tool for filtering information in complex systems. Proc Natl Acad Sci U S A 102(30):10421–10426
Wagner EH, Austin BT, Von Korff M (1996) Organizing care for patients with chronic illness. Milbank Q 74:511–544
Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A (2001) Improving chronic illness care: translating evidence into action. Health Affairs 20(6):64–78. Nov–Dec 2001
Wagner EH (1998) Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1(1):2–4. Aug–Sept 1998
Wasserman S, Faust K (1994) Social network analysis. Methods and applications. Cambridge University Press, Cambridge
Weber EP, Khademian AM (2008) Wicked problems, knowledge challenges, and collaborative capacity builders in network settings. Public Adm Rev 68(2):334–349
World Health Organization and Alzheimer's Disease International (2012) Dementia: a public health priority

Part IV
Care Process: Precaution

Chapter 9
Design and Planning of Organ Transplantation Networks

Sahar Ahmadvand and Mir Saman Pishvaee

9.1 Introduction to Organ Transplantation Network Management

9.1.1 Importance and Drivers

Organ transplantation has become one of the most successful and popular remedy methods due to technological and medical advances over the past 50 years. It is considered as the most cost-effective treatment for end-stage renal diseases and the only treatment for end stage failure of organs such as liver, lung and heart (Friele 2013). Despite all the advances in medication and increasing number of implantations in most of the countries, there still exist thousands of patients waiting for organs at the end of each year and many of them die before they are able to receive an organ transplant. In the United States of America 122,544 candidates were on the waiting list as of September 2015. It is notable that every 10 min one candidate is added to the waiting list and an average of 22 people die every day from the lack of organs (OPTN 2016).

The demand for organs always exceeds the supply and it arouses the most confusing dilemma in organ transplantation field which no country has found a solution for. Therefore, organs are considered as valuable scarce national resources, and efficient policies and strategies should be planned to manage them through the supply network. The efficient and collaborative management of organ transplantation network (OTN) activities including donation, procurement, preservation,

S. Ahmadvand • M.S. Pishvaee (✉)
School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran
e-mail: s_ahmadvand@ind.iust.ac.ir; pishvaee@iust.ac.ir

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_9

transportation and implantation of organs intended for transplantation is obtained through designing a well-organized OTN as well as planning the relevant operations all through the OTN.

9.1.2 Definitions and Scope

"Transplantation is the transfer of human cells, tissues or organs from a donor to a recipient with the aim of restoring function(s) in the body" (WHO 2014). Solid transplantable organs are "liver, heart, kidney, lung, pancreas and intestine". Most studies on OTN are devoted to kidney, liver and heart since they are the most needed and transplanted organs in the world (UNOS 2016).

In this chapter organ transplantation refers to the human solid organs transfer. The transplantation process involves a donor and a recipient. There are two types of donors: (1) deceased donor, and (2) living donor. Deceased donor is "an individual from whom at least one organ is recovered for transplant after declaration of death" and living donor is "a living individual from whom at least one organ is recovered for transplantation". Living donors can donate one-half of a paired organ set (e.g., one kidney) or a portion of an organ (e.g., a portion of the liver or a lobe of the lung) (UNOS 2016). Recipient is "the human being into whom allogeneic human organs, tissues or cells are transplanted" (WHO 2014).

Organ transplantation comprises two main phases: (1) procurement, and (2) surgery. Logistics management and design of OTN revolves around the procurement phase, while the surgery phase is restricted to medicine. The procurement phase is further decomposed into: (1) matchmaking, (2) transport routes planning, and (3) medical teams scheduling (Fuzzati 2005).

Several international and national organizations are established in different countries to administer these activities. Eurotransplant in Europe, united network for organ sharing (UNOS) in the United States of America and Iranian network for organ procurement (INOP) in Iran are examples of international and national organizations respectively. UNOS manages the transplantation activities through local organ procurement organizations (OPOs). These nonprofit units are in charge of identification and evaluation of brain death cases, obtainment of donor consent and procurement of organs (UNOS 2016). In the United States of America, the whole country is categorized into 11 UNOS regions, and further into 58 local donor service areas. Each of these local areas is under the jurisdiction of one OPO. Once an organ is available, the local OPO's waitlist is searched for a suitable recipient. The incompatible patients are screened out, and others are ordered according to some organ-specific criteria. If no suitable patient is found in the local area, the search takes place regionally, and thereafter nationally. This is the hierarchical allocation method. The way transplantation activities are managed and accomplished by INOP is by far similar to the way they are managed by UNOS. INOP is in charge of central management and allocation of organs and manages transplantation activities through local Organ Procurement Units (OPUs) set up in each university of medical

science. These units which are almost similar to OPOs are in control of identification and evaluation of brain death cases, obtainment of donor consent and procurement of organs. Each unit has a chief transplant coordinator who is responsible for making connection with INOP. The allocation of organs in INOP is done based on a hierarchical conduction like in UNOS (Kazemeyni et al. 2004). Iran which is a smaller country than USA is divided into 15 regions, and further into 30 local donor service areas. When an organ is procured, the matchmaking process starts with searching local OPU's waitlist and in case of finding no suitable recipient it continues by searching the upper level waitlists (MOHME 2016). Eurotransplant allocates organs based on another allocation method named centralized method. In this method the search and matchmaking process takes place centrally in the national/international-level waitlist (Çay 2012).

The management of organ transplantation process is a sophisticated task due to many restrictions such as time constraint. Deceased donors are kept artificially alive and this status cannot last for a long time. In addition once an organ is harvested from donor body, it should be implanted into the recipient's body within the organ's maximum allowable Cold Ischemia Time (CIT) (4–6 h for heart and lung, 8–12 h for liver, 24–36 h for kidney) (UNOS 2016); otherwise it will be wasted. CIT is the time during which the organ can survive in the absence of blood perfusion. It starts after organ harvesting operation and ends up by implantation of organ to the recipient body. It is desirable to minimize the CIT of organ in order to obtain better post-transplantation outcomes. Thus in many papers studying long-term decision problems, one objective is to minimize the transplantation time and consequently CIT (Beliën et al. 2013; Zahiri et al. 2014a).

It has been observed that availability of organs always does not lead to an efficient and on-time transplant. Sometimes an organ is donated, preserved and a suitable recipient is chosen for it, but the unavailability of medical staff and long transportation time to the recipient's TC may result in the recipient being discarded and the process of finding a suitable recipient being restarted. This causes the wastage of time and sometimes the wastage of organ (Fuzzati 2005). This mismanagement exacerbates the organ shortage. Hence, efficient coordination among medical staff, time management and planning for transportation routes in OPOs are essential. This arouses the long term decision problems of designing an efficient network, finding the best locations for network facilities (e.g., TCs, OPOs and shipping agents) and designing the best configuration of regions for hierarchical allocation systems.

Another important issue in the optimization of OTN that arises from ever-increasing gap between supply of and demand for organs is the design and evaluation of fair and efficient distribution and allocation policies. Organ allocation systems need to be as equitable as possible while being medically efficient because of the high value of organs as scarce resources. The terms equity and efficiency refer to two main conflicting objectives of allocation systems. The concept of distributive justice which implies the fair distribution of resources among all members of a society has relevance for organ allocation in view of the fact that organs are scarce public resources. Two main distributive justice criteria including (1) equal access and (2) maximum benefit are usually applied in this context.

Equal access aims to provide all patients with equal access to organs and decreases biases based on sex, race, socioeconomic status and geographic distance from donated organ. Waiting time (i.e., offering organs to patients on a 'first come first transplanted' basis) and age (i.e., ranking patients on a 'youngest to oldest' basis) are instances of equal access criteria (Benjamin 1988).

Maximum benefit criteria maximize the number of successful organ transplants. Medical need (i.e., ranking patients for receiving transplantable organs according to their urgency level) and probable success of a transplant (i.e., giving transplantation priority to the patient who is expected to live the longest after transplantation) are examples of maximum benefit criteria. These criteria ensure that by giving priority to the sickest patients and those who are likely to have better and longer post-transplantation lives, organ wastage is minimized and therefore allocation system is more efficient.

There are some debates around the biases, favoritism and unfair distribution of organs which may originate from the subjective nature of the maximum benefit criteria and polemicise against these criteria (Childress 2001). In fact, there is a conflict between efficiency and equity criteria which emerges the need for studying tradeoffs between them when designing allocation policies and algorithms. Hence, the conflicting nature of efficiency and equity is another source of difficulty faced by decision makers. Many mathematical programming models are proposed to study the tradeoff between efficiency and equity in literature (Zenios 2004). Taking into account equity and efficiency measures is important not only in designing allocation systems as short/mid-term decision problems but in designing the OTN and related long-term decision problems.

Some researchers have studied the problem of designing a transplantation network, through location-allocation of facilities and minimizing the time elapsed between harvesting and transplantation surgeries (Beliën et al. 2013), while others have focused on the problem of designing a regional configuration for organ allocation hierarchy that maximizes the society's medical efficiency and equity (Demirci et al. 2012).

Among all the activities involved in procurement phase, organ allocation is the most important and controversial short-term decision problem since it determines who lives and who dies. The need for an equitable and medically efficient allocation system has urged researchers to focus on designing and evaluating allocation policies (Zenios 2004).

The remainder of this chapter is prepared as follows: Section 9.2 contains a systematic review and classification of existing literature. In Sect. 9.3 two useful mathematical programming models are selected and discussed in detail. A credibility-based fuzzy approach is proposed to solve the first model under uncertainty. In Sect. 9.4 some related case studies are presented to help readers internalize the application of mathematical programming models in real-life environment. In Sect. 9.5 literature gaps and future research directions are offered to provide the enthusiastic readers with a variety of potential novel issues and subjects in OTN management.

9.2 Literature Review

The existing OTN planning problems can be categorized with respect to different planning horizons. They range from long-term decision problems to mid-term and short-term decision problems. Before reviewing papers in detail, we propose a hierarchical planning matrix for OTN. Two levels of decisions, i.e., long-term and short-term decisions, can be considered in different stages of OTN. Figure 9.1 illustrates the OTN hierarchical planning matrix with the organ transplantation process stages as horizontal dimension and the planning levels as vertical dimension. As previously mentioned organ transplantation activities comprise organ donation, organ allocation, organ/recipient/medical staff transportation, surgery and recovery/rehabilitation. The strategic-level decisions are related to the structure of OTN. In the OTN design, location-allocation of facilities (e.g., OPOs/OPUs, TCs and shipping agents), regional configuration design for hierarchical allocation system and contracting problems are considered as strategic-level decisions, while organ allocation, transportation of organs, recipients and medical staff, and personnel scheduling are the examples of tactical/operational-level decisions. It is worth mentioning that the location-allocation of OTN has obtained much attention in the literature; however, less emphasis is placed on the other relevant decision making problems, e.g., transportation planning, personnel planning and scheduling and planning for bed capacities in TCs.

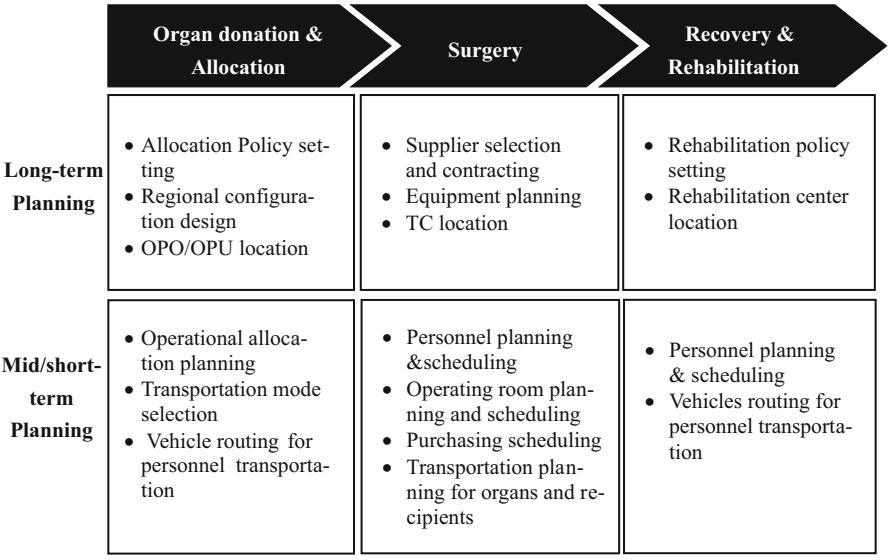

Fig. 9.1 OTN hierarchical planning matrix

To review the existing mathematical programming models in organ transplantation literature in a systematic way, firstly, we classify the papers from the planning level point of view into two main classes: (1) papers studying long-term (strategic) decision problems, and (2) papers studying short-term (operational) decision problems. Secondly, a coding system is developed to classify the existing works in the relevant literature.

Organ Transplantation Network Strategic Planning Models

As discussed before, strategic planning of organ transplantation deals with the long-term decisions such as location-allocation of network facilities and regional configuration design of hierarchical allocation systems in order to maximize society's benefit and minimize the whole network's costs (Zahiri et al. 2014b).

In this territory, Stahl et al. (2005) proposed an integer programming model to determine an optimal configuration of transplant regions while maximizing efficiency and geographic equity. They maximized a weighted combination of geographic equity and the number of transplants in each region as a measure of efficiency. The concerned problem was constrained by the number of regions for liver hierarchical allocation system in USA. Bruni et al. (2006) studied the problem of optimal design of multi-organ transplantation network in Italy. They used a mixed integer linear programming model to find optimal location of OTN facilities including OPOs, hospitals and TCs. The minimization of total travel time was considered as the objective function.

Kong et al. (2010) considered the problem of designing regional configuration for liver hierarchical allocation system in USA as a set partitioning problem. They developed a mixed integer programming model with a branch and price perspective to solve the problem. They aimed to maximize the number of intra-regional transplantations. Since the proposed model was NP-hard a heuristic approach was developed as an efficient solution method to solve the model. Çay (2012) studied the organ transplantation logistics network focusing on the Turkey case. She offered an integer linear programming model constrained by three types of equity constraints to formulate the regional coordination centers location-allocation problem while maximizing intra-regional organ flow. Furthermore, Demirci et al. (2012) studied the problem of redesigning configuration of regions for liver allocation system in USA, considering tradeoffs between a measure of geographic equity and efficiency. They used a branch and price approach to solve the proposed integer programming model and provided a heuristic method to approximate the frontier of Pareto-efficient solutions with respect to the objectives.

Beliën et al. (2013) presented a MILP model to tackle the Belgium OTN. They addressed the problem of optimally locating TCs, while taking into account the minimization of total weighted time as objective function. Zahiri et al. (2014b) proposed a robust possibilistic programming model for the location-allocation of organ transplantation supply chain in Iran under uncertainty. The objective function of the proposed MILP model was to minimize the weighted total costs and maximize the network's efficiency consequently. Furthermore, Zahiri et al. (2014a) studied the case of Iran from another point of view. They offered a MINLP model for

designing an organ transplant transportation network under uncertainty. The model has two objective functions including minimization of cost and time. They presented a fuzzy multi-objective programming approach to solve small problems and two meta-heuristic algorithms to solve large-sized problems.

Organ Transplantation Network Operational Planning Models

Since the most important part of organ transplantation decisions in short-time level is allocation of organs, the majority of operational planning research works are devoted to study organ allocation and distribution problems, notably kidney and liver allocation problems. Organ allocation systems and related policies undergo rapid changes because of the controversial and questionable nature of allocation and distribution decisions and existence of different alternatives (Alagoz et al. 2009). So far various mathematical programming models including analytical and computational models are developed and implemented to study organ allocation problem from different points of views. Another stream of works contains simulation models aiming at evaluation and comparison of various allocation policies. A number of mathematical programming models consider the problem from patients' (i.e., who may wish to accept or reject an organ offer) perspective while others concern the central decision makers' (e.g., UNOS policy makers) perspective. From the patients perspective the objective is to maximize patients' benefit and from central decision makers point of view the objective is to maximize society's benefit. In another classification a category of works tackles the development of an allocation policy whereas the other one deals with performance evaluation of allocation systems (Roth et al. 2003). Kidney has obtained much attention compared to other organs in mathematical modeling papers.

David and Yechiali (1985) provided a mathematical model in order to help patients decide whether to accept a living organ donation. They applied their model for the case of kidney. In another work by David and Yechiali (1990) the allocation of multiple organs to multiple recipients was formulated as a sequential stochastic assignment model. The objective of this problem was to find an allocation policy which maximizes various optimality criteria. Furthermore David and Yechiali (1995) considered a sequential matching problem for the assignment of multiple living donors to multiple waiting recipients. They assumed that an assignment gains a reward R if the donor and recipient match, and gains a reward r less than R if they do not. They discounted future rewards at a rate $0 \leq \alpha \leq 1$. The proposed approach was able to determine the optimal policies which maximize total discounted reward.

Ahn and Hornberger (1996) designed a decision making model to take into consideration the patients' preferences for acceptance of an offered organ. They developed a Markov Decision Process model to maximize patients' total expected quality adjusted life years employing expected 1-year graft survival rate as an index for measuring kidney acceptability.

Zenios (1999) developed a queuing model with reneging to represent the kidney transplant waiting list. He classified patients and organs according to their immunological and demographic characteristics and developed closed form expressions to identify the main factors underlying the performance of the transplant waiting list.

Zenios et al. (2000) proposed a dynamic resource allocation model to find an optimal allocation policy. The concerned objective was to maximize the quality adjusted life expectancy and minimize two measures of inequity. A dynamic index policy was obtained through approximate analysis of the optimal control problem and compared with the UNOS policies using a large-scale simulation model. Howard (2002) addressed the problem of acceptance or rejection of an offered liver as an optimal stopping problem. The problem of optimally timing a living-donor liver transplant was formulated as a Markov Decision Process by Alagoz et al. (2004). The objective was to maximize the patient's total expected reward, namely quality adjusted life years. Su and Zenios (2005) offered a sequential stochastic model to study the effect of patients' preferences on the allocation of kidneys. Their approach aimed at finding an optimal allocation policy that maximizes total expected reward for patients. In another paper by Su and Zenios (2006) the kidney allocation problem was framed as a sequential stochastic problem which considered both patients' and society's perspective. The goal was to determine an allocation policy which maximizes patients' total expected reward and society's equity. Alagoz et al. (2007) provided a Markov Decision Process model to help patients decide whether to accept a cadaveric liver organ. They considered patient health status and organ quality as the state of the process.

Last but not the least, Bertsimas et al. (2013) proposed a mechanism for estimating the optimal weights of scoring rules involved in the point systems for kidney allocation. Their method generally designed a point system, which was based on the selected scoring components and maximized medical efficiency, e.g., life year gains from transplant (LYFT), while simultaneously enforcing selected fairness constraints. This mechanism enables decision makers to dynamically change the scoring rules and evaluate the implementation results by means of a simulation model which is provided and used by OPTN. They input historical data, alternative score components and fairness constraints into a MINLP model and obtained optimal scoring rules using a linear regression.

Systematic Classification of the Literature

To the best of our knowledge, two different classifications are presented for organ allocation papers so far. Zenios (2004) presented a taxonomy of kidney allocation mathematical programming models. In a later work by Alagoz et al. (2009) organ allocation models were categorized according to organ type and decision makers' perspective. We take a broader view for classification of the literature by considering organ transplantation strategic-level models as well as organ allocation models in our taxonomy. According to the classification criteria offered in Table 9.1, we classify and tabulate most of the papers in existing literature as represented in Table 9.2.

Table 9.1 Classification criteria and related abbreviations

Decision perspective		**Equity objectives**	
Patient	p	Geographic equity	GE
Society	S	Likelihood of transplantation	LOT
Studied organ	Mean waiting time	MWT	
Kidney	K	**Uncertainty modeling approach**	
Liver	L	Fuzzy programming	FP
Multi organ	MO	Robust programming	RP
Modeling approach	**Outputs**		
Integer linear programming	ILP	Optimal allocation policy	OLP
Markov decision process	MDP	Acceptability of organ for patient	AOP
Mixed integer non-linear programming	MINLP	Closed form expressions	CFE
Integer programming	IP	Dynamic index policy	DIP
Sequential stochastic programming	SSA	Optimal timing for transplant	OT
Optimal stopping	OS	Mechanism design model	MDM
Fluid modeling	FM	Optimal regions	OR
Queuing model	QM	Optimal weights for scoring rules	OWSR
Mixed integer linear programming	MILP	Location-allocation of OPO	LAOPO
Integer non-linear programming	INLP	Location-allocation of TC	LATC
Efficiency objectives	Region design	RD	
Expected discounted reward	EDR	**Solution method**	
Total expected discounted reward	TEDR	Dynamic programming	DP
Total discounted reward	TDR	Exact	E
Average discounted reward	ADR	Linear regression	LR
Quality adjusted life expectancy	QALE	Policy improvement algorithm	PIA
Life years from transplant	LYFT	Markov process tools	MPT
Patient survival rate	PSR	Value iteration algorithm	VIA
Travel distance	TD	Policy iteration algorithm	PITA
Intra regional transplants	IRT	Achievable region	AR
Total transplantation process time	TTPT	Heuristic	H
Total cost	TCS	Meta heuristic	MH
Total weighted travel time	TWTT		
Quality adjusted life years	QALY		

Table 9.2 Taxonomy of literature

Article	Decision Perspective	Studied Organ	Modeling approach	Objective	Uncertainty modeling approach	Outputs	Solution method
Strategic-level planning models							
Stahl et al. (2005)	S	L	INLP	IRT, GE	–	LAOPO, RD	E
Bruni et al. (2006)	S	MO	MILP	TD	–	LATC	E
Kong et al. (2010)	S	L	MILP	IRT	–	LAOPO, RD	E, H
Demirci et al. (2012)	S	L	IP	PSR, GE	–	RD	H
Çay (2012)	S	MO	ILP	IRT	–	LAOPO	E
Beliën et al. (2013)	S	MO	MILP	TWTT	–	LAOPO,LATC	E
Zahiri et al. (2014b)	S	MO	MILP	TC	RP	LAOPO, LATC	E
Zahiri et al. (2014a)	S	MO	MILP	TTPT, TCS	FP	LAOPO	MH
Operational-level planning models							
David & Yechiali (1985)	P	K	OP	EDR	–	OAP	NA
David & Yechiali (1990)	S	MO	SSA	ADR	_	OLP	NA
David & Yechiali (1995)	S	K	SSA	TDR	–	OAP	DP
Ahn & Hornberger (1996)	P	K	MDP	QALE	–	AOP	E
Zenios (1999)	S	K	QM	NA	–	CFE	NA
Zenios et al. (2000)	S	K	FM	QALY, LOT, MWT	–	DIP	PIA
Howard (2002)	P	L	OS	TER	–	AOP	MPT
Alagoz et al. (2004)	P	L	MDP	QALE	–	OT	VITA
Su & Zenios (2005)	S, P	K	SSA	TER	–	OAP	DP
Su & Zenios (2006)	S, P	K	SSA	QLAY, GE	–	MDM	AR
Alagoz et al. (2007)	P	L	MDP	TEDR	–	AOP	PITA
Bertsimas et al. (2013)	S	K	MINLP	LYFT	–	OWSR	E,LR

9.3 Selected Mathematical Programming Models

In this section we discuss two selected organ transplantation planning models in detail. The first model is chosen from the category of strategic-level planning models tainted with uncertainties and the second one is an operational-level planning model. We propose a credibility-based fuzzy programming approach to handle uncertainties in the first model. The notations used here are the same as those used in the original papers to facilitate the reference to those papers.

9.3.1 A Multi-period Location-Allocation Model for Organ Transplant Centers

The location-allocation problem of organ TCs in Iran is studied by Zahiri et al. (2014b). They presented a novel robust possibilistic approach to cope with uncertainty of the input data. However, in this section we restrict our attention to the deterministic formulation rather than robust possibilistic formulation and further we present a credibility-based fuzzy programming approach in Sect. 9.3.2 to deal with uncertainty of input data.

Zahiri et al. (2014b) have considered an international OTN, since foreign donors and recipients are involved in the problem. Figure 9.2 is an illustration of the studied international supply chain and the existing relations among its different facilities. According to this illustration when an organ becomes available (1), the shipping team is sent to the donor hospital to get the information and required samples (2). Thereafter, the samples are delivered to the TC, where various medical analysis and tests take place (3). The shipping team is sent back to the hospital then (4). At this point the procedure of searching for the most suitable recipient and TC starts. The procured organ from domestic donor is sent to the selected TC, where the transplantation happens (5). The selected recipient is also transported to TC (6). In case of a foreign donor or recipient the donated organ is sent to the airport to arrive at TC for operation ($5'$).

The problem is formulated as a multi-period location-allocation model for facilities including hospitals, TCs and shipping agents. Given a number of potential places for different facilities and demands for various organs in different recipient zones, the model aims at making the best strategic decisions (e.g., finding the optimal locations of TCs, hospitals and shipping agents and the optimal assignment of flows between them) and designing an efficient network, while minimizing the total cost. The following notations are used to address the mathematical formulation of the proposed model. It is notable that uncertain parameters and variables are shown by a tilde ($\sim$) on.

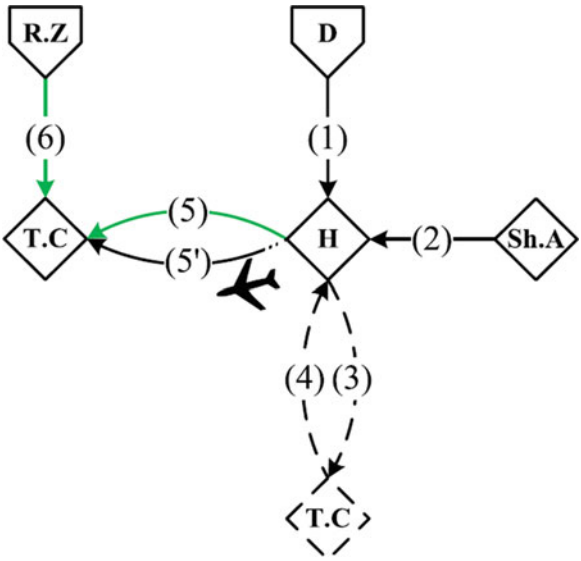

Fig. 9.2 The studied OTN (Adopted from Zahiri et al. 2014b)

Sets

I	Set of candidate hospital locations, $i \in I$
K	Set of candidate TC locations,$k \in K$
V	Set of shipping agent locations, $v \in V$
H	Set of recipient zones, $h \in H$
O	Set of organ types, $o \in O$
A	Set of airports, $a \in A$
T	Set of time periods, $t \in T$
E	Set of merged candidate sites between hospitals and TCs $e \in E, E \subset I, E \subset K$

Parameters

$\tilde{c}_i$	Fixed cost of establishing a hospital at location i
c'_k	Fixed cost of establishing a TC at location k
$\tilde{r}_{io}$	Cost of harvesting organ o at hospital i
r'_{ko}	Cost of equipping TC k for organ type o
Sd_i^t	Number of domestic donors at hospital i at time period t
Sa_i^t	Number of foreign donors at hospital i at time period t
M_{io}^t	Number of organ type o extracted from an individual donor at hospital i at time period t
Xd_i^t	Domestic donors' arrival rate at hospital i at time period t
Xa_i^t	Foreign donors' arrival rate at hospital i at time period t

$\tilde{D}_{ho}^{t}$	Demand of recipient zone h for organ type o at time period t
$\tilde{c}_{vi}^{(v\to i)}$	Cost of contract between shipping agent v and hospital i
$\tilde{c}_{ik}^{(i\to k)}$	Transportation cost for moving information and samples from hospital i to TC k
$c'^{(i\to k)}_{ik}$	Transportation cost for moving an organ from hospital i to TC k
$\tilde{c}_{hk}^{(h\to k)}$	Transportation cost for moving an individual from recipient zone h to TC k
$\tilde{c}_{ak}^{(a\to k)}$	Transportation cost for moving an organ from airport a to TC k
$\tilde{t}_{iko}^{t}$	Travelling time for moving organ type o from hospital i to TC k at time period t
t'_o	Allowable CIT for organ type o
M'	Penalty cost for unsatisfied demand
f'_e	Saving cost for merging facilities in candidate location e
λ	A weight for long-term planning level costs

According to the aforementioned notations the model formulation is presented as follows.

$$\begin{aligned}\min\ z = {} & \lambda\left(\sum_{i}^{I} c_i z_i + \sum_{k}^{K} c'_k z'_k + \sum_{o}^{O}\sum_{k}^{K} r'_{ko} y'_{ko} - \sum_{o}^{O}\sum_{e}^{E} f'_e y_{eo} y'_{eo}\right) \\ & + \sum_{t}^{T}\sum_{o}^{O}\sum_{a}^{A}\sum_{k}^{K}\sum_{i}^{I} r_{io}\left(X'^{(i\to k)t}_{iko} + X^{(i\to a)t}_{iao}\right) + \sum_{t}^{T}\sum_{i}^{I}\sum_{v}^{V} \tilde{c}_{vi}^{(v\to i)} X_{vi}^{(v\to i)t} \\ & + \sum_{t}^{T}\sum_{o}^{O}\sum_{k}^{K}\sum_{i}^{I} c_{ik}^{(i\to k)} X_{iko}^{(i\to k)t} + \sum_{t}^{T}\sum_{o}^{O}\sum_{k}^{K}\sum_{i}^{I} c'^{(i\to k)}_{ik} X'^{(i\to k)t}_{iko} \\ & + \sum_{t}^{T}\sum_{o}^{O}\sum_{a}^{A}\sum_{k}^{K} c_{ak}^{(a\to k)} X_{ako}^{(a\to k)t} + \sum_{t}^{T}\sum_{o}^{O}\sum_{k}^{K}\sum_{h}^{H} \tilde{c}_{hk}^{(h\to k)} X_{hko}^{(h\to k)t} + \sum_{h}^{H} \widetilde{M'} B_h\end{aligned} \tag{9.1}$$

s.t.

$$y_{io} \le z_i, \forall i, o, \tag{9.2}$$

$$y'_{ko} \le z'_k, \forall k, o, \tag{9.3}$$

$$\sum_{i}^{I} y_{io} \ge 1, \forall o, \tag{9.4}$$

$$\sum_{k}^{K} j'_{ko} \ge 1, \forall o, \tag{9.5}$$

Variables

z_i	$\begin{cases} 1 & \text{If a hospital is located at candidate location i} \\ 0 & \text{otherwise} \end{cases}$
z'_k	$\begin{cases} 1 & \text{If a TC is located at candidate location k} \\ 0 & \text{otherwise} \end{cases}$
y_{io}	$\begin{cases} 1 & \text{If a hospital capable of donating organ o is located at candidate location i} \\ 0 & \text{otherwise} \end{cases}$
y'_{ko}	$\begin{cases} 1 & \text{If a TC equipped for organ o is located at candidate location k} \\ 0 & \text{otherwise} \end{cases}$
w_v^t	$\begin{cases} 1 & \text{If a shipping agent v is selected at time period t} \\ 0 & \text{otherwise} \end{cases}$
$X_{vi}^{(v\to i)t}$	$\begin{cases} 1 & \text{If a hospital at location i is covered by shipping agent v at time period t} \\ 0 & \text{otherwise} \end{cases}$
$X_{iko}^{(i\to k)t}$	Flow of information and samples of organ type o from hospital i to TC k at time period t
$X'^{(i\to k)t}_{iko}$	Flow of organ type o from hospital i to TC k at time period t
$X_{iao}^{(i\to a)t}$	Flow of organ type o from hospital i to airport a at time period t
$X_{ako}^{(a\to k)t}$	Flow of organ type o from airport a to TC k at time period t
$X_{hko}^{(h\to k)t}$	Flow of recipients from recipient zone h to TC k at time period t
$sd_{io}^{\prime t}$	Available number of organ type o at hospital i procured from domestic donors
$sa_{io}^{\prime t}$	Available number of organ type o at hospital i procured from foreign donors
$I_{io}^{\prime t}$	Inventory level of organ type o at hospital i at time period t
B_h	Unsatisfied demand for organs in recipient zone h
W	Number of available shipping agents at each time period

$$\sum_{v}^{V} w_v^t = W, \forall t, \tag{9.6}$$

$$X_{vi}^{(v\to i)t} \leq w_v^t, \forall i, v, t, \tag{9.7}$$

$$y_{io} \leq \sum_{v}^{V} X_{vi}^{(v \to i)t}, \forall i, o, t, \tag{9.8}$$

$$w_v^t \leq \sum_{i}^{I} X_{vi}^{(v \to i)t}, \forall v, t, \tag{9.9}$$

$$\sum_{v}^{V} X_{vi}^{(v \to i)t} \leq 1, \forall i, t, \tag{9.10}$$

$$\sum_{k}^{K} X_{iko}^{(i \to k)t} = \left(sd_{io}'^t + sa_{io}'^t\right) y_{io}, \forall i, o, t, \tag{9.11}$$

$$X_{iko}^{(i \to k)t} \leq \left(sd_{io}'^t + sa_{io}'^t\right) y_{ko}', \forall i, k, o, t, \tag{9.12}$$

$$X_{iko}'^{(i \to k)t} = 0 \left| \tilde{t}_{iko}^t > t_o' \right., \forall i, k, o, t, \tag{9.13}$$

$$X_{iko}'^{(i \to k)t} \leq sd_{io}'^t y_{ko}', \forall i, k, o, t, \tag{9.14}$$

$$\sum_{k}^{K} X_{iko}'^{(i \to k)t} \leq sd_{io}'^t y_{io}, \forall i, o, t, \tag{9.15}$$

$$\sum_{a}^{A} X_{iao}^{(i \to a)t} \leq sa_{io}'^t y_{io}, \forall i, o, t, \tag{9.16}$$

$$\sum_{a}^{A} X_{ako}^{(a \to k)t} \leq \sum_{i}^{I} sa_{io}'^t y_{ko}', \forall k, o, t, \tag{9.17}$$

$$\sum_{k}^{K} X_{ako}^{(a \to k)t} = \sum_{i}^{I} X_{iao}^{(i \to a)t}, \forall a, o, t, \tag{9.18}$$

$$sd_{io}'^t = sd_i^t M_{io}^t X d_i^t, \forall i, o, t, \tag{9.19}$$

$$sa'^t{}_{io} = sa_i^t M_{io}^t X a_i^t, \forall i, o, t, \tag{9.20}$$

$$\sum_{h}^{H} X_{hko}^{(h\to k)t} = \sum_{i}^{I} X_{iko}^{\prime(i\to k)t} + \sum_{a}^{A} X_{ako}^{(a\to k)t}, \forall k, o, t, \tag{9.21}$$

$$\sum_{k}^{K} X_{hko}^{(h\to k)t} + B_h \geq \tilde{D}_{ho}^{t}, \forall h, o, t, \tag{9.22}$$

$$I_{io}^{t} = I_{io}^{t-1} + sd'^{t}_{io} + sa'^{t}_{io} - \left(\sum_{k=1}^{K} X_{iko}^{\prime(i\to k)t} + \sum_{a=1}^{A} X_{iao}^{(i\to a)t} \right), \forall i, o, t, \left| I_{io}^{0} = 0 \right. \tag{9.23}$$

$$z_i, z'_k, y_{io}, y'_{ko}, w_v^t, X_{vi}^{(v\to i)t} \in \{0, 1\}, \forall i, k, o, t, v, \tag{9.24}$$

$$X_{iko}^{(i\to k)t}, X_{iko}^{\prime(i\to k)t}, X_{iao}^{(i\to a)t}, X_{ako}^{(a\to k)t}, X_{hko}^{(h\to k)t} \geq 0, \text{integer} \forall i, k, o, h, a, t \tag{9.25}$$

$$sd_{io}^{\prime t}, sa_{io}^{\prime t}, I_{io}^{t}, W \geq 0, \text{integer}, \forall i, o, t. \tag{9.26}$$

Objective function (9.1) minimizes weighted total costs including long-term planning costs and periodic costs, namely harvesting operation costs, transportation costs, cost of unsatisfied demands and saving costs due to merging hospitals and TCs at possible candidate locations. Constraints (9.2) and (9.3) respectively ensure that a hospital or TC is equipped for organ type o if and only if it is opened. Constraints (9.4) and (9.5) ensure the existence of at least one hospital and TC equipped for each organ type. Constraint set (9.6) defines the exact number of shipping agents at each time period. Constraint set (9.7) assures that a hospital can be under a shipping agent's service domain only if the shipping agent is selected. Constraint set (9.8) guarantees that at least one shipping agent is assigned to each hospital. Constraint set (9.9) indicates that a shipping agent is selected only if it is assigned to at least one hospital, in other words this constraint minimizes the number of unused shipping agents. Constraint set (9.10) makes sure that each hospital gets service from at most one shipping agent. Constraint sets (9.11) and (9.12) ensure that the various flows among hospitals and TCs are feasible if and only if the related facilities are opened. Constraint set (9.13) checks that the travelling time for a procured organ does not exceed the maximum allowable CIT of the organ. Constraint sets (9.14) and (9.15) ensure that the organ flow from a hospital to a TC is feasible only if the hospital and TC are opened and the amount of the flow is limited by the supply of the organ in the hospital. Constraint set (9.16) ensures that the organ flow from a hospital to the airports is possible only if the hospital is equipped for harvesting that specific organ type. Constraint set (9.17) indicates the necessity of opening a specific TC for the feasibility of flows from airports to that TC. Constraint set (9.18) assures that total flows of organs procured from foreign donors from hospitals to an airport equals

total flows of organs from that airport to TCs. Constraint sets (9.19) and (9.20) estimate the total number of procured domestic and foreign organs for each type in each hospital at each time period, respectively. Constraint (9.21) indicates that total flows from recipient zones to a TC should be equal to total incoming flows from hospitals and airports to the TC. Constraint set (9.22) checks the demand satisfaction of each recipient zone. Constraint set (9.23) indicates the inventory level of organ supplies of each type, in each hospital and at each time period. Constraint sets (9.24, 9.25, and 9.26) declare the types of decision variables.

It is notable that in the aforementioned formulation, the objective function and some of the constraints are nonlinear due to the multiplication of two or more variables. The authors of the paper applied linearization methods to convert the model into its linear counterpart. Enthusiastic readers may consult the original paper for more details on linearization part.

9.3.2 A Credibility-Based Fuzzy Programming Approach to Multi-period Location-Allocation of Organ Transplant Centers Under Uncertainty

Zahiri et al. (2014b) studied the problem of organ transplant supply chain location allocation under uncertainty. They proposed a robust possibilistic programming approach to deal with uncertainties of the problem and considered parameters to be trapezoidal fuzzy numbers. In this section we develop and present a credibility-based fuzzy programming approach to the mentioned problem rather than introducing their approach. For further details on robust possibilistic programming approach we refer enthusiastic readers to the original paper.

The credibility approach converts understudy fuzzy model into a credibility-based fuzzy multi-period location allocation model by using credibility measures of fuzzy constraints. It is shown that for trapezoidal fuzzy membership functions the fuzzy credibility-based multi-period location allocation of transplant centers becomes a linear programming model.

Credibility-based chance constrained programming is one of the efficient and most advantageous methods of possibilistic programming since it uses credibility measure which is a self-dual measure unlike possibility and necessity measures (Liu and Liu 2002). In other words once the credibility measure equals one it is obvious that the fuzzy event certainly takes place and when the credibility measure equals zero the fuzzy event certainly does not take place theoretically. But when the possibility measure of a fuzzy event equals one there is a chance that the event does not take place. Similarly a necessity measure of zero does not assure that the event won't take place. The proposed credibility-based fuzzy multi-period location allocation of organ transplant centers is formulated as follows

$$\min u \tag{9.27}$$

$$cr\left\{\begin{array}{l}\lambda\left(\sum\limits_{i}^{I}c_i z_i+\sum\limits_{k}^{K}c'_k z'_k+\sum\limits_{o}^{O}\sum\limits_{k}^{K}r'_{ko}y'_{ko}-\sum\limits_{o}^{O}\sum\limits_{e}^{E}f'_e y_{eo}y'_{eo}\right)\\ +\sum\limits_{t}^{T}\sum\limits_{o}^{O}\sum\limits_{a}^{A}\sum\limits_{k}^{K}\sum\limits_{i}^{I}r_{io}\left(X'^{(i\to k)t}_{iko}+X^{(i\to a)t}_{iao}\right)+\sum\limits_{t}^{T}\sum\limits_{i}^{I}\sum\limits_{v}^{V}\tilde{c}^{(v\to i)}_{vi}X^{(v\to i)t}_{vi}\\ +\sum\limits_{t}^{T}\sum\limits_{o}^{O}\sum\limits_{k}^{K}\sum\limits_{i}^{I}c^{(i\to k)}_{ik}X^{(i\to k)t}_{iko}+\sum\limits_{t}^{T}\sum\limits_{o}^{O}\sum\limits_{k}^{K}\sum\limits_{i}^{I}c'^{(i\to k)}_{ik}X'^{(i\to k)t}_{iko}+\\ \sum\limits_{t}^{T}\sum\limits_{o}^{O}\sum\limits_{a}^{A}\sum\limits_{k}^{K}c^{(a\to k)}_{ak}X^{(a\to k)t}_{ako}+\sum\limits_{t}^{T}\sum\limits_{o}^{O}\sum\limits_{k}^{K}\sum\limits_{h}^{H}\tilde{c}^{(h\to k)}_{hk}X^{(h\to k)t}_{hko}+\sum\limits_{h}^{H}\widetilde{M'}B_h\le u\end{array}\right\}\ge\alpha_0 \tag{9.28}$$

$$cr\left\{X'^{(i\to k)t}_{iko}=0\,|\tilde{t}^{t}_{iko}>t'_o\right\}\ge\alpha_1,\forall i,k,o,t, \tag{9.29}$$

$$cr\left\{sd'^{t}_{io}=sd^{t}_{i}M^{t}_{io}Xd^{t}_{i}\right\}\ge\alpha_2,\forall i,o,t, \tag{9.30}$$

$$cr\left\{sa'^{t}_{io}=sa^{t}_{i}M^{t}_{io}Xa^{t}_{i}\right\}\ge\alpha_3,\forall i,o,t, \tag{9.31}$$

$$cr\left\{\sum_{k}^{K}X^{(h\to k)t}_{hko}+B_h\ge\tilde{D}^{t}_{ho}\right\}\ge\alpha_4,\forall h,o,t, \tag{9.32}$$

Constraints (9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 9.10, 9.11, and 9.12), (9.14, 9.15, 9.16, 9.17, and 9.18), (9.21) and (9.23, 9.24, 9.25, and 9.26), where u is an auxiliary free variable and α_i is the known acceptable credibility level of each fuzzy constraint. Suppose that ξ is a trapezoidal fuzzy number $\xi=(\xi^{(1)},\xi^{(2)},\xi^{(3)},\xi^{(4)})$ and r is a real number, according to Liu (2004) and Zhu and Zhang (2009) we have the following equations for $\alpha\ge 0.5$.

$$cr\{\xi\le r\}\ge\alpha\leftrightarrow r\ge(2-2\alpha)\,\xi^{(3)}+(2\alpha-1)\,\xi^{(4)} \tag{9.33}$$

$$cr\{\xi\ge r\}\ge\alpha\leftrightarrow r\le(2\alpha-1)\,\xi^{(1)}+(2-2\alpha)\,\xi^{(2)} \tag{9.34}$$

The readers are referred to the original papers for further details and proofs. Equations 9.33 and 9.34 can directly be used for transforming fuzzy chance constraints to their deterministic counterparts. To convert fuzzy chance constraint $cr\{\xi=r\}\ge\alpha$ to its deterministic counterpart we use the approach proposed by Liu (2004). In this approach for $a\ge 0.5$ we have following equation

$$cr\{\xi\approx r\}\ge\alpha\leftrightarrow\left\{\begin{array}{l}r\ge\left(2\left(\frac{\alpha}{2}\right)-1\right)\xi^{(1)}+\left(2-2\left(\frac{\alpha}{2}\right)\right)\xi^{(2)}\\ r\le\left(2-2\left(\frac{\alpha}{2}\right)\right)\xi^{(3)}+\left(1-2\left(\frac{\alpha}{2}\right)\right)\xi^{(4)}\end{array}\right. \tag{9.35}$$

According to Eqs. (9.33 and 9.35) the suggested credibility-based fuzzy multi-period location allocation model can be deffuzified and reformulated as Eqs. (9.36 and 9.39).

$$\min u \tag{9.36}$$

$$(2-2\alpha_0)\left[\begin{array}{l}\lambda\left(\sum_{i}^{I} c^{(3)}{}_i z_i + \sum_{k}^{K} c'^{(3)}_k z'_k + \sum_{o}^{O}\sum_{k}^{K} r'^{(3)}_{ko} y'_{ko} - \sum_{o}^{O}\sum_{e}^{E} f'^{(2)}_e y_{eo} y'_{eo}\right) \\ + \sum_{t}^{T}\sum_{o}^{O}\sum_{a}^{A}\sum_{k}^{K}\sum_{i}^{I} r_{io}{}^{(3)}\left(X'^{(i\to k)t}_{iko} + X^{(i\to a)t}_{iao}\right) + \sum_{t}^{T}\sum_{i}^{I}\sum_{v}^{V} c^{(3)}{}^{(v\to i)}_{vi} X^{(v\to i)t}_{vi} \\ + \sum_{t}^{T}\sum_{o}^{O}\sum_{k}^{K}\sum_{i}^{I} c^{(3)}{}^{(i\to k)}_{ik} X^{(i\to k)t}_{iko} + \sum_{t}^{T}\sum_{o}^{O}\sum_{k}^{K}\sum_{i}^{I} c'^{(3)}{}^{(i\to k)}_{ik} X'^{(i\to k)t}_{iko} + \\ \sum_{t}^{T}\sum_{o}^{O}\sum_{a}^{A}\sum_{k}^{K} c^{(3)}{}^{(a\to k)}_{ak} X^{(a\to k)t}_{ako} + \sum_{t}^{T}\sum_{o}^{O}\sum_{k}^{K}\sum_{h}^{H} c^{(3)}{}^{(h\to k)}_{hk} X^{(h\to k)t}_{hko} + \sum_{h}^{H} M'^{(3)} B_h\end{array}\right]$$

$$+(2\alpha_0-1)\left[\begin{array}{l}\lambda\left(\sum_{i}^{I} c^{(4)}{}_i z_i + \sum_{k}^{K} c'^{(4)}_k z'_k + \sum_{o}^{O}\sum_{k}^{K} r'^{(4)}_{ko} y'_{ko} - \sum_{o}^{O}\sum_{e}^{E} f'^{(1)}_e y_{eo} y'_{eo}\right) \\ + \sum_{t}^{T}\sum_{o}^{O}\sum_{a}^{A}\sum_{k}^{K}\sum_{i}^{I} r_{io}{}^{(4)}\left(X'^{(i\to k)t}_{iko} + X^{(i\to a)t}_{iao}\right) + \sum_{t}^{T}\sum_{i}^{I}\sum_{v}^{V} c^{(4)}{}^{(v\to i)}_{vi} X^{(v\to i)t}_{vi} \\ + \sum_{t}^{T}\sum_{o}^{O}\sum_{k}^{K}\sum_{i}^{I} c^{(4)}{}^{(i\to k)}_{ik} X^{(i\to k)t}_{iko} + \sum_{t}^{T}\sum_{o}^{O}\sum_{k}^{K}\sum_{i}^{I} c'^{(4)}{}^{(i\to k)}_{ik} X'^{(i\to k)t}_{iko} + \\ \sum_{t}^{T}\sum_{o}^{O}\sum_{a}^{A}\sum_{k}^{K} c^{(4)}{}^{(a\to k)}_{ak} X^{(a\to k)t}_{ako} + \sum_{t}^{T}\sum_{o}^{O}\sum_{k}^{K}\sum_{h}^{H} c^{(4)}{}^{(h\to k)}_{hk} X^{(h\to k)t}_{hko} + \sum_{h}^{H} M'^{(4)} B_h\end{array}\right] \geq u \tag{9.37}$$

$$X'^{(i\to k)t}_{iko} = 0 \left| (2\alpha_1 - 1)\, t^{(1)t}_{iko} + (2-2\alpha_1)\, t^{(2)t}_{iko} > t'_o \right., \forall i,k,o,t, \tag{9.38}$$

$$sd'^{t}_{io} \geq \left[\left(2\left(\frac{\alpha_2}{2}\right)-1\right) sd^{(1)t}_{i} + \left(2-2\left(\frac{\alpha_2}{2}\right)\right) sd^{(2)t}_{i}\right] M^{t}_{io} Xd^{t}_{i}, \forall i,o,t, \tag{9.39}$$

$$sd'^{t}_{io} \leq \left[\left(2-2\left(\frac{\alpha_2}{2}\right)\right) sd^{(3)t}_{i} + \left(1-2\left(\frac{\alpha_2}{2}\right)\right) sd^{(4)t}_{i}\right] M^{t}_{io} Xd^{t}_{i}, \forall i,o,t, \tag{9.40}$$

$$sa'^{t}_{io} \geq \left[\left(2\left(\frac{\alpha_3}{2}\right)-1\right) sa^{(1)t}_{i} + \left(2-2\left(\frac{\alpha_3}{2}\right)\right) sa^{(2)t}_{i}\right] M^{t}_{io} Xa^{t}_{i}, \forall i,o,t, \tag{9.41}$$

$$sa'^{t}_{io} \leq \left[\left(2-2\left(\frac{\alpha_3}{2}\right)\right) sa^{(3)t}_{i} + \left(1-2\left(\frac{\alpha_3}{2}\right)\right) sa^{(4)t}_{i}\right] M^{t}_{io} Xa^{t}_{i}, \forall i,o,t, \tag{9.42}$$

$$\sum_{k}^{K} X^{(h\to k)t}_{hko} + B_h \geq (2-2\alpha_4)\, D^{(3)t}_{ho} + (2\alpha_4 - 1)\, D^{(4)t}_{ho} \forall h,o,t, \tag{9.43}$$

constraints (9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 9.10, 9.11, and 9.12), (9.14, 9.15, 9.16, 9.17, and 9.18), (9.21) and (9.23, 9.24, 9.25, and 9.26).

The latter formulation is completely deffuzified using credibility measure but is not linear because of multiplication of variables in objective function and some constraints. After linearization of nonlinear terms and constraints according to methods used by Zahiri et al. (2014b) the credibility-based fuzzy multi-period location allocation of transplant centers model becomes a linear programming model as we claimed.

9.3.3 *A Scalable, Data-Driven Method for Designing Fair and Efficient Kidney Allocation Policies*

Bertsimas et al. (2013) proposed a scalable, data-driven method for designing fair and efficient kidney allocation policies based on point systems being used in USA. A point system ranks patients according to some specified criteria, namely medical status and waiting time. Their method provides central decision makers with the flexibility of choosing their arbitrary criteria and fairness constraints to construct various allocation policies. Once the desired criteria and fairness constraints are chosen, the method by choosing optimal score weights designs an allocation policy which aims at maximizing a measure of medical efficiency, i.e., life year gains from transplant, while satisfying specified fairness constraints. This method is also a means of evaluation and comparison of different scoring rules.

OPTN kidney transplantation committee is responsible for the distribution and allocation of deceased donor kidneys according to developed allocation policies. Allocation policies need to fit into economical, medical, ethical, institutional and other societal factors. According to OPTN an allocation policy must

- Use organs as efficient as possible and avoid organ wastage,
- Rank patients according to medical judgments,
- Balance the medical efficiency (e.g., life year gains from transplant) and equity (e.g., waiting time).

So far, all implemented allocation policies have been designed based on point systems or scoring rules. Scoring rule, ranks patients according to a calculated score, namely *Kidney Allocation Score* (KAS). A scoring rule is a weighted sum of some score components. A score component is a function of one or some of the patient's or organ's characteristics. Consider the following notations.

O:	Set of available organs, $o \in O$
P:	Set of patients on the waiting list, $p \in P$
J:	Set of score components, $j \in J$

When an organ becomes available, the KAS for each medically compatible patient-organ pair is calculated as Eq. 9.44.

$$KAS\,(p,o) = \sum_{j=1}^{|J|} w_j f_j\,(p,o) \tag{9.44}$$

The calculated value for jth score component is indicated by $f_j(p,o)$ and the jth score weight is shown by w_j.

Bertsimas et al. (2013) introduced following items as examples of criteria based on which score components are designed:

- Tissue matching or HLA matches, i.e., the number of HLAs patient p and organ o have in common;
- Age of patient p and organ o donor;
- Waiting time of patient p;
- Dialysis time (DT) for patient p;
- Blood group of donor and patient;
- Expected post-transplant survival of patient p receiving organ o;
- Expected survival of patient p on the waitlist;
- Life year gains from transplant for patient p receiving organ o;
- Calculated panel reactive antibody (CPRA), denoting the sensitization of the patient;
- Donor profile index (DPI), denoting the quality of the donated organ.

Scoring rules developed by KTC usually comprise 3–5 score components which are either linear or nonlinear functions of some of the aforementioned criteria. Assigning points per year on dialysis or per life years from transplant are the examples of such linear functions. Awarding points according to multiplication of DPI and LYFT or multiplication of DT and DPI to the corresponding patients are instances of nonlinear functions. Among all the 40 scoring rules proposed by KTC, the following scoring rule seemed to be paramount as of 2008:

$$\begin{aligned} KAS\,(p,o) &= 0.8LYFT\,(p,o) \times (1 - DPI(o)) \\ &+ 0.8DT(p) \times DPI(o) + 0.2DT(p) + 0.04CPRA(p). \end{aligned} \tag{9.45}$$

The ever-changing nature of the allocation policies results from the question of whether there exists another scoring rule comprising the same components, which performs more equitably and efficiently. Many simulation models are designed to evaluate the outcomes of different policies in order to respond to this challenging question. The method developed by Bertsimas and his co-workers functions as a helpful tool in this area and means to respond to the above question.

This method inputs selected score components, fairness constraints and historical data and outputs weights associated with the score components, while maximizing medical efficiency.

Methodology

For any procured organ, set C denoting the set of eligible patient-organ pairs can be built based on physiological and medical characteristics of the patients and organs.

$$C = \{(p, o) : \text{Patient p is eligible to receive organ o}\} \tag{9.46}$$

Score components and LYFT associated with each eligible pair can be calculated thereafter.

Binary decision variable $x_{(p,o)}$ for each eligible pair of patient-organ $(p,\ o)$ is defined as:

$$x_{(p,o)} = \begin{cases} 1 \text{ If organ o is allocated to patient p} \\ 0 \qquad\qquad\qquad \text{otherwise} \end{cases} \tag{9.47}$$

An admissible constraint for this methodology is a linear constraint with respect to the variable$x_{(p,o)}$. The fairness constraints used by OPTN ensure that at least a specific percentage of total number of transplants is assigned to each specific group of patients. For instance, a constraint imposing a lower bound L on the number of transplants for a particular group of patients $G \subset P$, is formulated as

$$\sum_{p \in G} \sum_{o:(p,o) \in C} x_{(p,o)} \geq L \tag{9.48}$$

Enforcing a lower bound on the number of organs allocated to the group of blood type O patients is an example of such constraints. Bertsimas and his co-workers express the fairness constraints by $Ax \leq b$ for some matrix A, and vector b. They consider all registered patients and procured organs over a fixed time period and suppose all patients accept the organ offers. Based on such assumptions they formulated the problem of allocating organs to patients with the objective function of maximizing efficiency, i.e., life year gains from transplant, subject to desired fairness constraints $Ax \leq b$, as a linear optimization problem:

$$\begin{aligned} &\max \sum_{(p,o) \in C} LYFT_{(p,o)} x_{(p,o)} \\ &s.t. \\ &\sum_{o;(p,o) \in C} x_{(p,o)} \leq 1, \qquad \forall \text{p} \\ &\sum_{p;(p,o) \in C} x_{(p,o)} \leq 1, \qquad \forall \text{o} \\ &Ax \leq b \\ &x \geq 0. \end{aligned} \tag{9.49}$$

Assume that y is the vector of dual multipliers associated with the constraints $Ax \leq b$, by linear optimization duality, Model (9.49) can be reformulated as Model (9.50).

$$\begin{aligned} \max & \sum_{(p,o)\in C} LYFT_{(p,o)} x_{(p,o)} - y^T Ax + y^T b \\ & s.t. \\ & \sum_{o;(p,o)\in C} x_{(p,o)} \le 1, \qquad \forall \mathrm{p} \\ & \sum_{p;(p,o)\in C} x_{(p,o)} \le 1, \qquad \forall \mathrm{o} \\ & x \ge 0. \end{aligned} \tag{9.50}$$

The objective function of Model (9.50) can be rewritten as $c^T x + y^T b$, using cost vector c as shown in Eq. (9.51).

$$c_{(p,o)} = LYFT_{(p,o)} - \left(y^T A\right)_{(p,o)}, \forall\, (p,o) \in C. \tag{9.51}$$

The aim is finding an allocation policy which solves the above matching problem online. One may think of ranking patients according to the above $c_{(p,o)}$ coefficients, but these coefficients are artificial and it is preferred to use selected score components. Hence, admissible score weights w can be obtained from the solution to the optimization problem (9.52).

$$\begin{aligned} \min & \sum_{(p,o)\in C} \left(c_{(p,o)} - w_0 - \sum_{j=1}^{|J|} w_j f_{j,(p,o)} \right)^2 \\ & s.t. \\ & w \in S. \end{aligned} \tag{9.52}$$

Given a set of score components J, data for linear constraints (A, b), historical data including (1) set of eligible patient-organ pairs C, (2) estimated LYFT for each pair and (3) calculated values of score components for each pair $f_{j,(p,o)}$, we can summarize the proposed methodology for finding the score weights as following steps:

Step1: Solve problem (9.49);
Step2: Find vector of optimal dual multipliers y associated with the constraint sets $Ax \le b$;
Step3: Compute$c_{(p,o)} = LYFT_{(p,o)} - (y^T A)_{(p,o)}, \forall\, (p,o) \in C$;
Step4: Use linear regression (9.53) to estimate the score weights$w_0, w_1, \ldots, w_{|J|}$.

$$c_{(p,o)} \approx w_0 + w_1 f_{1,(p,o)} + \cdots + w_{|J|} f_{|J|,(p,o)} \tag{9.53}$$

9.4 Case Study

In this section we move from theory to practice. We review two selected case studies from recent literature works. The first case study belongs to the category of strategic-level planning problems and is the application of a multi-objective location-allocation model proposed by Zahiri et al. (2014a) to the Iranian OTN. The second case study is adopted from the paper by Bertsimas et al. (2013). The authors of the second paper utilize their proposed method to construct three scoring rule based allocation policies respective to three different case studies according to organ transplantation legislations in USA. We include their first case study in this section and refer interested readers to Bertsimas et al. (2013) for more case studies.

Case Study 1: Organ Transplantation Network Design in Iran
Zahiri et al. (2014a) proposed a multi-period location-allocation model for the design of an OTN. They formulated the problem as a bi-objective mixed-integer programming model which aims at minimizing total cost including transportation cost and fixed facility opening cost and procurement time including different travelling times, waiting time in TCs and operation times in hospitals. They applied their model to design a cost-effective and efficient OTN for Iran. The considered assumptions are as follows:

- The potential locations of hospitals and TCs are determined based on experts' estimations. As shown in Fig. 9.3, there are 25 and 4 candidate locations for hospitals and TCs respectively.
- Twenty-two provinces of Iran are assumed as recipient zones due to their high population density and immensity.
- As Iran transplantation supply chain network does not include shipping agents, these components are screened out from the problem. Noteworthy, air and ground are two transportation modes considered in this study.
- The input data including fixed opening and equipping costs, transportation costs, travelling times, operation times, supplies and demands are tainted with uncertainty.

The developed chance constrained programming model is transformed into two models: (1) lower approximation model (LAM) and (2) upper approximation model (UAM). An optimistic-pessimistic parameter ξ, $(0 \leq \xi \leq 1)$ is used to indicate the combined perspective of decision-makers. A hybrid solution approach is utilized to solve the bi-objective model and the corresponding optimal results of LAM and UAM models are depicted in Figs. 9.4 and 9.5, respectively. According to illustrated results the number of hospitals to be established in UAM model is greater than that of LAM but the number of TCs to be opened is equal in both models. Table 9.3 represents the optimal interval value for first objective function. The cost imposed by the current inefficient network to Iran's health system is 1.27 times greater.

Case Study 2: Design of a Scoring Rule Based Kidney Allocation Policy
In this case study a scoring rule-based allocation policy is designed using the methodology proposed by Bertsimas et al. (2013). This allocation policy is intended to have the same fairness properties and score components as the dominant policy

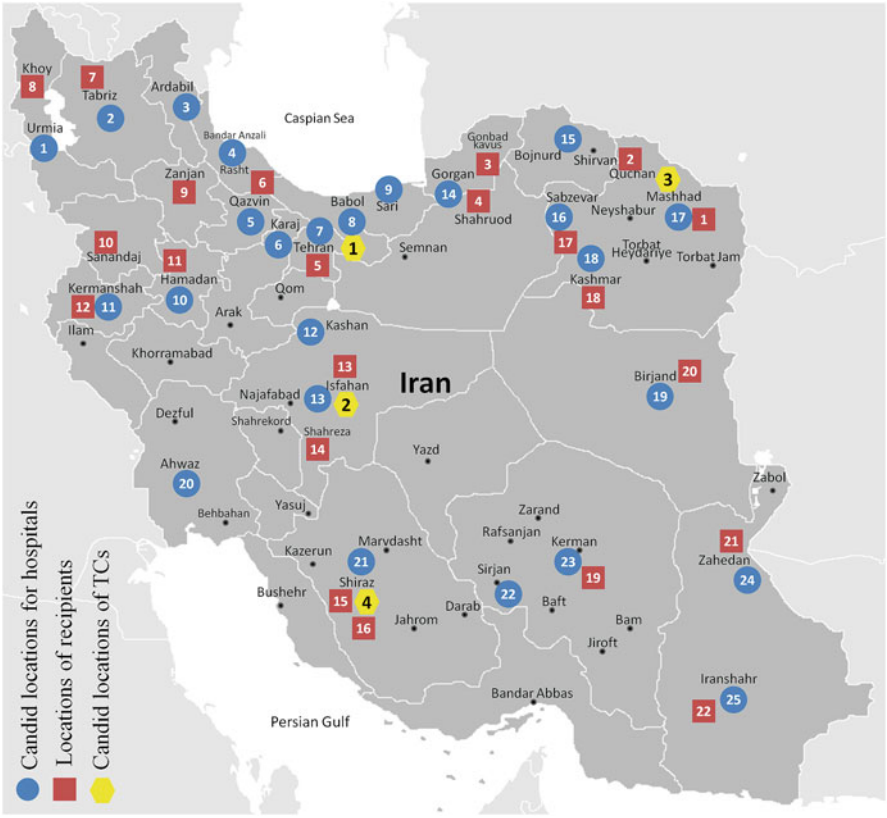

Fig. 9.3 Candidate locations of hospitals and TCs and indicators of recipient zones (Adopted from Zahiri et al. 2014a)

used by KTC. According to this methodology following inputs are needed to construct the policy:

- Historical data: The data of 2008 adopted from the Scientific Registry of Transplant Recipients (SRTR) data base, is divided into two 6 months. The first 6 months of data is used as training data and the remaining part of data is used for evaluation of the developed policy.
- Fairness constraints: As previously mentioned this allocation policy is intended to have the same fairness attributes as the dominant KTC policy. In order to ensure this condition, the KTC policy needs to be simulated over the training data set and the distribution of recipients across different patient groups (e.g., race, age, blood type, disease type, dialysis time, CPRA level groups) should be calculated and recorded as percentages. The lower bounds for fairness constraints are obtained from these statistical data. For instance, the recorded percentage of distribution of recipients among patients of different levels of DT for the KTC

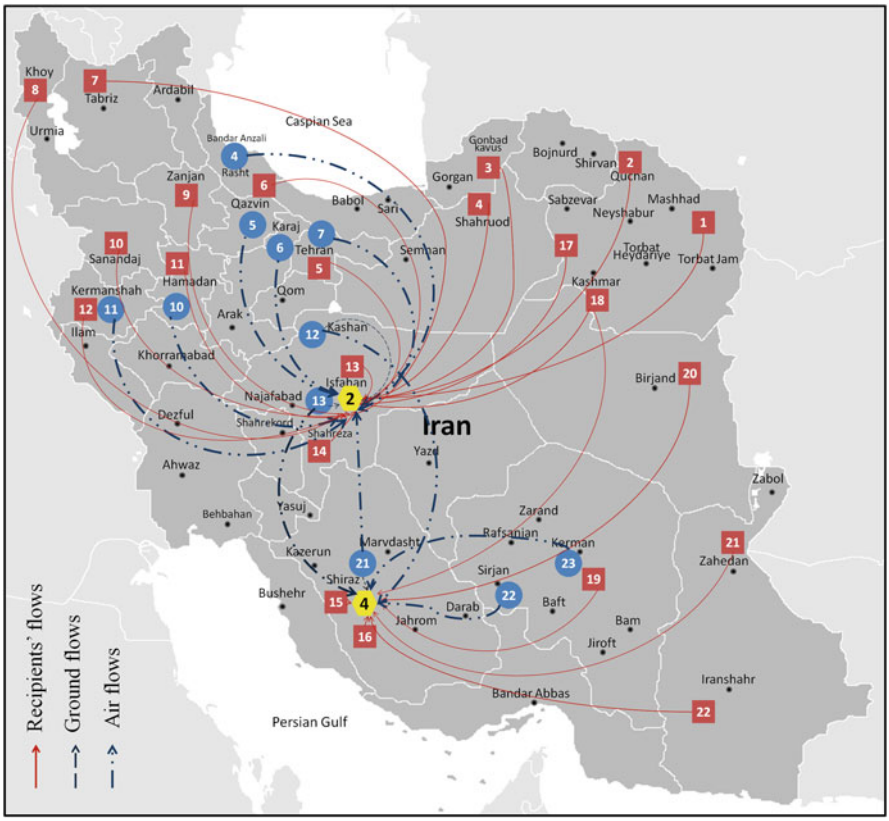

Fig. 9.4 LAM optimum network design and flows between facilities (Adopted from Zahiri et al. 2014a)

policy is as follows: 54% of the recipients have spent less than 5 years on dialysis, 29.5% from 5 to 10 years, 11.1% from 10 to 15 years and 5.4% more than 15 years. The related fairness constraints to be used in the model are expressed below:

$$\begin{aligned}
&\sum_{p:0\le DT(p)\le 5}\ \sum_{o:(p,c)\in C} x_{(p,o)} \ge \frac{54}{100} \sum_{(p,c)\in C} x_{(p,o)} \\
&\sum_{p:5\le DT(p)\le 10}\ \sum_{o:(p,c)\in C} x_{(p,o)} \ge \frac{29.5}{100} \sum_{(p,c)\in C} x_{(p,o)} \\
&\sum_{p:10\le DT(p)\le 15}\ \sum_{o:(p,c)\in C} x_{(p,o)} \ge \frac{11.1}{100} \sum_{(p,c)\in C} x_{(p,o)} \\
&\sum_{p:15\le DT(p)}\ \sum_{o:(p,c)\in C} x_{(p,o)} \ge \frac{5.4}{100} \sum_{(p,c)\in C} x_{(p,o)}
\end{aligned} \tag{9.54}$$

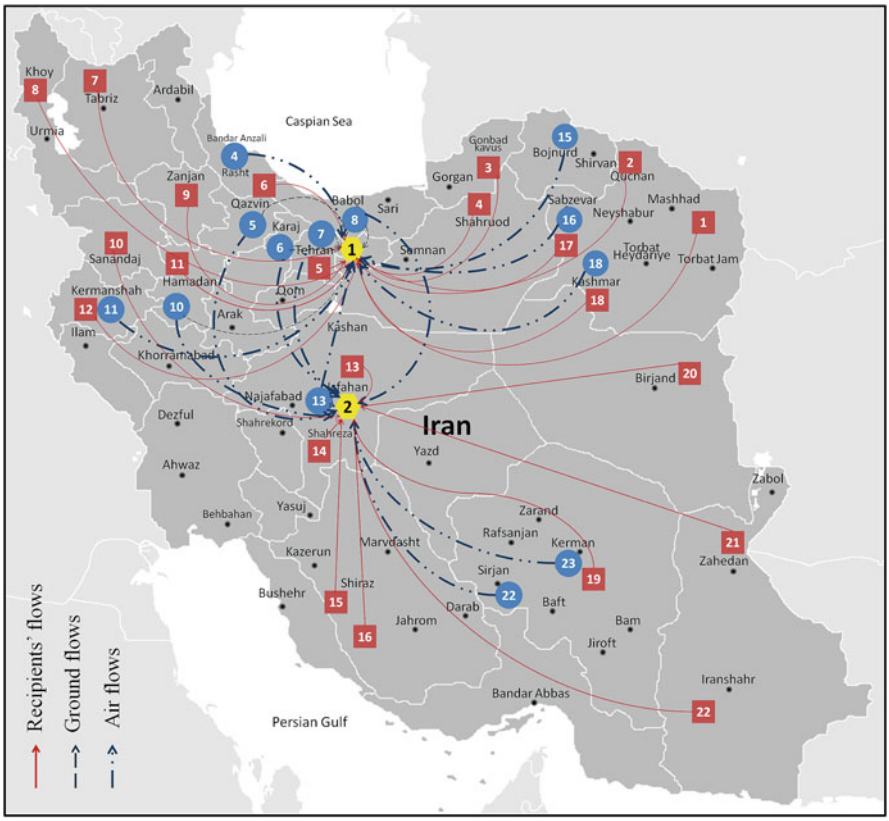

Fig. 9.5 UAM optimum network design and flows between facilities (Adopted from Zahiri et al. 2014a)

- Score components: All the score components considered by KTC (see Sect. 9.3.3) are put into the linear regression problem and finally four most significant ones are selected as in usual policies implemented by KTC: (1) LYFT, (2) a piece-wise linear function of DT, (3) CPRA and (4) a step-wise function of patient's age.

Taking into account all the aforementioned assumptions, the presented method results in finding optimal score weights, and following scoring rule which assigns KAS to each patient-organ pair consequently:

$$KAS\,(p,o) = LYFT\,(p,o) + g\,(DT(p)) + 0.08CPRA(p) + 0.5I\,(age(p) \geq 50) \tag{9.55}$$

Where I is the indicator function and

Table 9.3 Optimum interval of first objective function and current network's cost

	Optimum interval	Current network in Iran
Fixed cost	[97700,107100]	146700

Table 9.4 Simulation results for KTC policy and the policy presented in case study 1 over data related to the second half of 2008 and for 100 runs

	KTC policy	Case study 1
Number of transplantations (std)	5799 (23)	5807 (22)
Net life years from transplant (std)	34217 (195)	36890 (219)

$$g = \begin{cases} 0.65DT, & 0 \leq DT \leq 5, \\ DT - 1.75, & 5 \leq DT \leq 10, \\ 0.2DT + 6.25, & 10 \leq DT. \end{cases} \tag{9.56}$$

According to the above scoring rule, patients are awarded one point for each life year gain from transplant and 0.08 points for each CPRA level unit. 0.65 $\times$ DT points are assigned to patients spent 0–5 years on dialysis, DT-1.75 points goes to patients spent 5–10 years on dialysis, 0.2 $\times$ DT + 6.25 points are awarded to patients spent more than 10 years on dialysis. 0.5 points are assigned to patients older than 50 years.

As mentioned earlier, the simulation models are capable to evaluate and compare allocation policies. Bertsimas and his coworkers used the *Kidney-Pancreas Simulator Allocation Model* (KPSAM) which is developed and utilized by OPTN KTC for performance evaluation of allocation policies, to simulate the outcomes of their proposed allocation policies over different case studies.

To evaluate the policy from efficiency point of view, the average number of total transplantations and average total life year gains from transplant over 100 simulation runs are recorded and compared with those of KTC dominant policy. The evaluation from fairness point of view is done based on the percentage of recipients' distribution among various patients groups. Simulation results for efficiency evaluation are tabulated in Table 9.4. We refer the enthusiastic readers to Bertsimas et al. (2013) for fairness evaluation results.

Through this case study, Bertsimas and his coworkers could develop an allocation policy which is based on the same score components and fairness constraints used in KTC dominant policy and achieves a 7.8% increase in life year gains compared to the KTC policy.

9.5 Future Research Directions

Since mathematical programming models have a significant and explicit role in the efficient management of OTNs as one of the most critical branches of healthcare management sciences, there is an undeniable need in developing appropriate

mathematical models to cope with the optimization problems originating from the main dilemma of distributing and allocating the scarce organ resources efficiently and equitably. So far, many works have been devoted to study the OTN problems from various perspectives and planning level horizons. However, many aspects of the organ transplantation management have remained questioned and unstudied.

In many of the reviewed models, the authors restricted their definition of equity or efficiency to one or two measures and neglected other applicable components of equity/efficiency. Modifying existing models by considering other components of equity and efficiency, developing new models entailing multiple components and proposing a mathematical and clearer definition of equity and efficiency can be studied in future.

Organs by the reason of mismanagement in OTN are exposed to wastage. Since organs are scarce national resources, mathematical programming models can be employed to identify the detailed organ wastage reasons in order to minimize the total wastage.

A major part of papers studying organ transplantation issues has focused on kidney and liver as the two most transplanted organs around the world. Further, these models can be universalized and applied to other organ types.

As a pointer for future study, one can combine the two perspectives of decision making in organ allocation policy design, and assess the outcomes on efficiency and equity and find out whether the overall performance of the current allocation systems would improve if patients' preferences are taken into account during matchmaking process.

All the reviewed organ allocation models consider the medical factors underling allocation procedures and neglect the economic and logistics aspects of the allocation and distribution of organs. But the central decision makers may be willing to consider the economic and logistics aspects of the allocation and transportation problems as well as medical efficiency and equity, in designing an allocation system, since they face time and budget limitations in real environment. This issue arouses the need for combining the logistics planning perspective with common organ allocation problems, which could be a novel research area.

References

Ahn J-H, Hornberger JC (1996) Involving patients in the cadaveric kidney transplant allocation process: a decision-theoretic perspective. Manag Sci 42(5):629–641

Alagoz O, Maillart LM, Schaefer AJ, Roberts MS (2004) The optimal timing of living-donor liver transplantation. Manag Sci 50(10):1420–1430

Alagoz O, Maillart LM, Schaefer AJ, Roberts MS (2007) Determining the acceptance of cadaveric livers using an implicit model of the waiting list. Oper Res 55(1):24–36

Alagoz O, Schaefer AJ, Roberts MS (2009) Optimizing organ allocation and acceptance. In: Handbook of optimization in medicine, vol 26. Springer, pp 1–24

Beliën J, De Boeck L, Colpaert J, Devesse S, Van den Bossche F (2013) Optimizing the facility location design of organ transplant centers. Decis Support Syst 54(4):1568–1579

Benjamin M (1988) Medical ethics and economics of organ transplantation. Health Prog (Saint Louis, Mo) 69(2):47–52

Bertsimas D, Farias VF, Trichakis N (2013) Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. Oper Res 61(1):73–87
Bruni ME, Conforti D, Sicilia N, Trotta S (2006) A new organ transplantation location–allocation policy: a case study of Italy. Health Care Manag Sci 9(2):125–142
Çay P (2012) Organ transplantation logistics: case for Turkey. Bilkent University
Childress JF (2001) Putting patients first in organ allocation: an ethical analysis of the US debate. Camb Q Healthc Ethics 10(04):365–376
David I, Yechiali U (1985) A time-dependent stopping problem with application to live organ transplants. Oper Res 33(3):491–504
David I, Yechiali U (1990) Sequential assignment match processes with arrivals of candidates and offers. Probab Eng Informational Sci 4(04):413–430
David I, Yechiali U (1995) One-attribute sequential assignment match processes in discrete time. Oper Res 43(5):879–884
Demirci MC, Schaefer AJ, Romeijn HE, Roberts MS (2012) An exact method for balancing efficiency and equity in the liver allocation hierarchy. INFORMS J Comput 24(2):260–275
Friele R (2013) Study on the set-up of organ donation and transplantation in the EU Member States, uptake and impact of the EU Action Plan on Organ Donation and Transplantation (2009–2015)
Fuzzati R (2005) Organ transplantation management. Swiss Federal Institute of Technology Lausanne (EPFL) Technical Report, No: IC/2005/022
Howard DH (2002) Why do transplant surgeons turn down organs?: a model of the accept/reject decision. J Health Econ 21(6):957–969
Kazemeyni SM, Bagheri Chrime AR, Heidari AR (2004) Worldwide cadaveric organ donation systems (transplant organ procurement)
Kong N, Schaefer AJ, Hunsaker B, Roberts MS (2010) Maximizing the efficiency of the US liver allocation system through region design. Manag Sci 56(12):2111–2122
Liu B (2004) Uncertainty theory: an introduction to its axiomatic foundations. Springer, Berlin
Liu B, Liu Y-K (2002) Expected value of fuzzy variable and fuzzy expected value models. IEEE Trans Fuzzy Syst 10(4):445–450
MOHME (2016) Ministry of health and medical education (Iran). http://www.behdasht.gov.ir/
OPTN (2016) Organ procurement and transplantation network. https://optn.transplant.hrsa.gov/
Roth AE, Sonmez T, Unver MU (2003) Kidney exchange. National bureau of economic research, Cambridge, MA
Stahl JE, Kong N, Shechter SM, Schaefer AJ, Roberts MS (2005) A methodological framework for optimally reorganizing liver transplant regions. Med Decis Mak 25(1):35–46
Su X, Zenios SA (2005) Patient choice in kidney allocation: a sequential stochastic assignment model. Oper Res 53(3):443–455
Su X, Zenios SA (2006) Recipient choice can address the efficiency-equity trade-off in kidney transplantation: a mechanism design model. Manag Sci 52(11):1647–1660
UNOS (2016) United network for organ sharing. https://www.unos.org/data/
WHO (2014) World Health Organization. http://www.who.int/topics/transplantation/en/
Zahiri B, Tavakkoli-Moghaddam R, Mohammadi M, Jula P (2014a) Multi-objective design of an organ transplant network under uncertainty. Transp Res Part E: Logist Transp Rev 72:101–124
Zahiri B, Tavakkoli-Moghaddam R, Pishvaee MS (2014b) A robust possibilistic programming approach to multi-period location–allocation of organ transplant centers under uncertainty. Comput Ind Eng 74:139–148
Zenios SA (1999) Modeling the transplant waiting list: a queueing model with reneging. Queueing Syst 31(3–4):239–251
Zenios SA (2004) Models for kidney allocation. In: Operations research and health care, vol 70. Springer, pp 537–554
Zenios SA, Chertow GM, Wein LM (2000) Dynamic allocation of kidneys to candidates on the transplant waiting list. Oper Res 48(4):549–569
Zhu H, Zhang J (2009) A credibility-based fuzzy programming model for APP problem. In: Artificial intelligence and computational intelligence, 2009. AICI'09. International Conference on, 2009. IEEE, pp 455–459

Chapter 10
Blood Supply Chain Management and Future Research Opportunities

Ali Ekici, Okan Örsan Özener, and Elvin Çoban

10.1 Introduction

Blood is needed for several types of treatments including organ transplants, cancer and anemia treatments, and major surgeries such as open heart surgery. In many countries, people still die because of inadequate supply of blood products (World Health Organization 2016). Although there is a continuing research on blood substitutes, these attempts have not been successful yet. Therefore, the demand for blood and blood products is satisfied by the donation of eligible individuals. About 38% of the population is eligible for donation in the U.S. However, around 9.5 million people (3% of the population) donate blood in a year, and approximately 16 million units are collected (American Association of Blood Banks 2007).

Every year, around 5 million patients in the U.S. receive blood or blood products such as red blood cells, platelets or plasma via blood transfusions (American Red Cross 2011). In 2006, more than 30 million units of blood products were used in blood transfusions (American Association of Blood Banks 2007), and the usage rate of blood products is expected to rise due to increases in human life expectancy and advances in the medical procedures that require blood transfusions such as organ transplants and open heart surgery (American Association of Blood Banks 2015; Davey 2004). Although supply marginally balances aggregate demand currently, the growth rate of supply is significantly smaller than that of demand (American Association of Blood Banks 2007), due to aging of the regular donors and more stringent donor selection criteria (Katsaliaki 2008; Pitocco and Sexton 2005; Williams et al. 1997). Based on these factors, the demand for blood products is expected to exceed the supply within 20 years (Currie et al. 2004). Traditional

A. Ekici (✉) • O.Ö. Özener • E. Çoban
Department of Industrial Engineering, Ozyegin University, Istanbul, Turkey
e-mail: ali.ekici@ozyegin.edu.tr; orsan.ozener@ozyegin.edu.tr; elvin.coban@ozyegin.edu.tr

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_10

approaches to donor recruitment to keep up with the increase in demand have not been successful (Popovsky 2005). The combination of rising demand and shrinking donor pool forces blood donation organizations to be more efficient with the donors they have available (Kopach et al. 2008; Pitocco and Sexton 2005; Valbonesi et al. 2005).

Blood supply management is different from a typical supply chain problem in three aspects: (1) The number of donations, and thus, the amount of blood supply, changes over time. For example, in general, there are seasonal shortages because of low blood collections during the winter and summer months (American Association of Blood Banks 2015; Gilcher and McCombs 2005). Moreover, in Muslim countries the donations are very low during Ramadan month. (2) The whole blood and blood products have short life spans, and there has to be a certain amount of time between two consecutive donations of a donor to replace the donated blood. For example, whole blood can be stored for 35 days whereas red bloods cells have a shelf life of 42 days; platelets perish after 5 days, and plasma expires after 1 year. Furthermore, a whole blood donor has to wait for 56 days to make another donation or a donor can donate platelets every 7 days up to 24 times a year. (3) In order to extract platelets, one of the most important blood products, donated whole blood has to be processed within 6 h of donation.

Perishability of blood products and the deferral time between consecutive donations have two serious implications over a typical supply chain. First, having excess donations is not always better both because of donation/inventory related costs of outdated products and the ineligibility of the donor to donate for a certain amount of time. Second, inadequate number of donations results in postponed surgeries, untreated patients and thus lost lives. In 2011, 4.3% of the blood products processed were outdated, and in 3.3% of the hospitals, elective surgeries were postponed one or more days because of shortages (American Association of Blood Banks 2011). Given the short life span of blood products and the waiting time between consecutive donations by a single donor, the management of blood products supply is a highly dynamic problem. It is a great challenge for the blood collecting agencies to keep an adequate supply of blood products during normal times, not to mention in times of high demand due to natural or man-made disasters.

Since there is no substitute for blood products, the demand is satisfied only by donations. The most common type of blood donation is whole blood donation. After the blood is donated at a blood donation site, it can be stored as whole blood and used for patients during surgery or to replace blood lost. However, most of it is separated into components such as red blood cells, platelets and plasma by a process called *centrifugation* (see Fig. 10.1), and these products can be used to treat several medical conditions or diseases. In the U.S., the collection, storage, and processing of blood and its components are regulated by Food and Drug Administration (FDA) and the American Association of Blood Banks (AABB). According to these regulations, in order to extract platelets the collected whole blood must be separated into its components preferably within 6 h and not more than 18 h after collection. Therefore, the donated blood should be collected from the donation sites and delivered to a processing center within 6 h to separate it into its components.

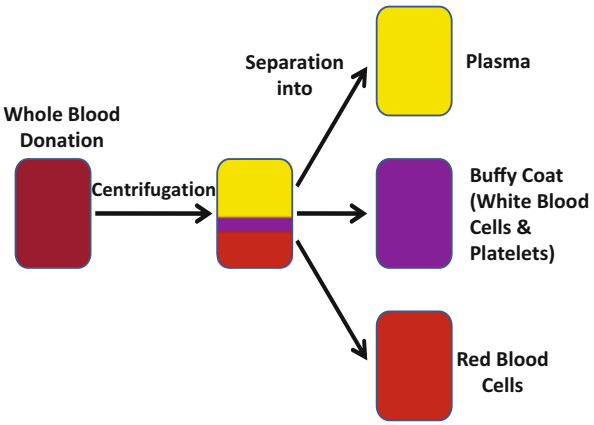

Fig. 10.1 Blood products extracted from a donated whole blood

Historically, whole blood was used in transfusions, but currently, only components of the blood are used. Plasma is used for burn and trauma patients, and for the treatment of clotting disorders. Red blood cells are the most commonly used component (more than 60% of units used are red blood cells (South Texas Blood & Tissue Center 2014)), and are needed for any patient requiring transfusion. They are mainly used for anemia treatment, surgery, treatment of blood disorders and for premature babies. Finally, platelets help control bleeding, and they are used in cancer treatments, organ transplants, and other surgeries to prevent massive blood loss. Out of one unit of whole blood donation (1 unit $\approx$ 450 ml), one unit of red blood cells, 0.5–1 unit of plasma and around 0.1–0.2 unit of platelets can be extracted. To illustrate the amount of demand, we provide some data about the approximate amount of blood products used in various transfusions. For instance, a bone marrow transplant patient needs 120 units of platelets and 20 units of red blood cells. A heart surgery patient needs 6 units of red blood cells and 6 units of platelets, and an automobile accident victim may need up to 50 units of red blood cells (Blood Bank of Alaska 2016).

In addition to whole blood donation, recent technological advances allow the donation of more than one blood product and/or more than one transfusable unit of each product without risking the donor safety based on the eligibility/characteristics of the donor. This method is called *multicomponent apheresis* (MCA) (also called *multicomponent collection* (MCC)) and is now frequently used by non-profit donation collection organizations and companies (Haemonetics Corporation 2008; Ridley 2009). To give an example for MCA, in red blood cell-plasma donation one unit of red blood cells and up to 3 units of plasma can be collected based on the donor's weight (Stanford Blood Center 2015), or in double red blood cell-double platelet donation, a donor can donate two units of red blood cells and two units of platelets (Valbonesi et al. 2005). Hence, by MCA, the donation process can be tailored based on the donor eligibility and the component(s) to be collected. The

deferral time changes based on the component collected. For example, a donor can donate whole blood or red blood cells every 56 days, double red blood cells every 112 days, and plasma every 28 days.

The main advantages of MCA donation are as follows: (1) More than one component and/or more than one unit can be collected in a single donation. This results in savings due to increased yield per donation, reduced donor time, reduced cost of extra bags, and reduced number of required tests before performing a transfusion which significantly reduces testing costs and time. Moreover, an additional processing step for separating the donated blood into its components is not required; the final product is ready after the donation. This reduces both the processing costs and the logistics costs due to transporting whole blood from the donation site to a processing center. The product(s) collected by MCA can be stored appropriately at the donation site until transported to the storage facility. (2) Donor utilization can be increased. For example, collecting two units of red blood cells and a unit of platelets from an eligible individual provides more transfusable units compared to a whole blood donation, which helps in efficiently managing the limited number of donors. Moreover, different customized eligibility criteria for each donation type, by the help of technically more developed devices, allows a donor to be eligible for a specific type of donation even if s/he is not eligible for whole blood donation (Mendez et al. 2007). (3) The patient is exposed to the blood of less number of donors which can minimize the infection risks. (4) It helps in stabilizing the type-distribution inventory and/or matching the demand and supply in a more cost-effective way by tailoring the donations according to the demand.

MCA donations provide an opportunity for improving the cost-effectiveness of the donation process and increasing the utilization of the donor pool (Bonomo et al. 2004; Popovsky 2005; Smith and Gilcher 2006; Vaquero 2006). Moreover, seasonal shortages during summer and winter months can be eliminated by tailoring the donations using MCA devices, and a more stable inventory level throughout the year can be maintained (Haemonetics Corporation 2008). To overcome the shortage of donors and the fluctuation of donation/usage patterns and to reduce soaring health care costs by improving donor utilization, the usage of MCA donations are expanding in many countries (Matthes 2002; Valbonesi et al. 2005). Economic analysis of MCA donations and exploring the potential benefits in developing customized schedules are very critical for an improved blood supply chain (Blanco 2002; Connelly and Pink 2002).

In this chapter, we first review the existing literature about blood supply management, and then discuss the research problems pertinent to blood collection operations including the collection/pickup scheduling and donation tailoring to meet the desired inventory levels and/or satisfy demand.

10.2 Existing Literature on Blood Supply Chain

Blood supply chain management has received much attention from researchers. Pierskalla (2004) provides an overview of the literature on blood supply management discussing the strategic level decisions such as assigning donor regions and transfusion centers to community blood centers, determining the number of community blood centers and locating these centers, and tactical and operational decisions such as determining target inventory levels, allocating inventory to hospitals, crossmatching, issuing, distribution and transshipment policies. More recently, Belien and Force (2012) provide a survey of the literature on the inventory and supply chain management of blood products, and classify the existing literature from different perspectives such as solution method, type of product studied and performance measures used.

Most of the studies in the literature focus on inventory management by considering the perishability of blood products. Goyal and Giri (2001), Karaesmen et al. (2011), Nahmias (1982) and Prastacos (1984) provide reviews of the literature on perishable inventory theory. In the perishable inventory models, the following factors are considered: (1) average inventory level, (2) average age of transfused blood, and (3) average number of shortages and wastage. The inventory models related to blood supply analyze the following aspects of the problem:

- *Target Inventory Levels and Order Policies:* Several models are developed to set inventory levels and determine ordering strategies for whole blood and blood products to achieve some balance between shortages and wastage at the regional and hospital blood bank levels (Brodheim et al. 1975, 1976; Chazan and Gal 1977; Cohen et al. 1979; Elston and Pickrel 1963; Fontaine et al. 2009; Friedman et al. 1982; Fries 1975; Haijema et al. 2007; Hesse et al. 1997; Hurlburt and Jones 1964; Jagannathan and Sen 1991; Jennings 1973; Kaspi and Perry 1983; Katz et al. 1983; Kendall 1980; Kopach et al. 2008; Ledman and Groh 1984; Nahmias 1976; Pereira 2005; Rytila and Spens 2006; Sirelson and Brodheim 1991; Vrat and Khan 1976). These include (1) simulation models (Abbott et al. 1978; Brodheim et al. 1976; Cohen and Pierskalla 1979; Jennings 1973), (2) Markov chain models (Brodheim et al. 1975; Chazan and Gal 1977; Jagannathan and Sen 1991; Pegels and Jelmert 1970), and (3) dynamic programming models (Fries 1975; Nahmias 1975; Nahmias and Pierskalla 1973; Prastacos 1981). Due to the stochastic nature of the blood supply chain, forecasts may not be reliable. Hence, researchers also develop models to handle demand uncertainty. For example, Van Dijk et al. (2009) study the platelet production and inventory management using a combined stochastic dynamic programming and simulation model. Haijema et al. (2009) extend their approach to include irregular production breaks like Easter and Christmas. Some authors differentiate between demand for "old" and "young" platelets and red blood cells, and develop 2-stage models to meet two different demand patterns (Goh et al. 1993; Haijema et al. 2007; Perry and Posner 1990).

- *Crossmatching Policies:* Crossmatching is a set of complex tests performed to ensure compatibility of patient's blood against units of blood from inventory. Crossmatched blood is reserved for the patient for a certain amount of time. Parameters such as the proportion of crossmatched blood that is actually transfused (transfusion-crossmatch ratio) and the number of days after which crossmatched blood is released if not transfused (crossmatch-release period) significantly affect the wastage. Several authors study the effect of double crossmatching (Dumas and Rabinowitz 1977; Katsaliaki 2008; Pegels and Jelmert 1970, 1971), and crossmatch release time (Cohen et al. 1983, 1979; Jagannathan and Sen 1991; Prastacos 1984; Rabinowitz 1973) on outdated units.
- *Issuing Policies:* Another area of policy control for blood inventory management is the issuing policy in terms of age sequence of units issued to meet demand. The most common issuing policies are FIFO (first-in-first-out, using oldest first) and LIFO (last-in-first-out, using freshest first). The effect of several issuing policies on the age of transfused blood, average inventory levels, wastage and shortage rates is investigated by several authors (Albright 1976; Cohen and Pierskalla 1975; Elston and Pickrel 1965; Pegels and Jelmert 1970; Pegels et al. 1977; Pierskalla and Roach 1972; Yen 1975). Pierskalla and Roach (1972) proves that FIFO minimizes cumulated outdates and shortages for some particular classes of perishable inventory problems.
- *Inventory Allocation and Distribution Policies:* These models focus on allocating blood from a regional blood center to the hospitals considering the shortages and outdates (Abbott et al. 1978; Cohen and Pierskalla 1979; Denesiuk et al. 2006; Elston and Pickrel 1965; Gregor et al. 1982; Jennings 1968; Pegels et al. 1977; Prastacos 1978, 1981; Sapountzis 1984, 1989; Yahnke et al. 1972; Yen 1975), and designing corresponding distribution schedules (Alshamrani et al. 2007; Federgruen et al. 1986; Hemmelmayr et al. 2009, 2010; Or and Pierskalla 1979). For example, Hemmelmayr et al. (2009) investigate the value of switching from current vendee managed inventory to vendor managed inventory in the presence of a deterministic usage rate. Due to uncertainty in the usage rates in practice, Hemmelmayr et al. (2010) extend this approach to handle stochastic product usage. Related to inventory allocation, another policy for rotating unused "still good" blood from one hospital to another one through regional blood center (which is called *rotation policy*) is also studied (Brodheim et al. 1975; Federgruen et al. 1986; Jennings 1973; Kendall 1980; Kendall and Lee 1980; Prastacos 1978, 1981). A successful real-life application of a blood rotation policy in Long Island is provided by Brodheim and Prastacos (1979a,b) and Prastacos and Brodheim (1979, 1980).

Other than these inventory models, several authors study (1) the regionalization and organization of the blood banking supply chain to manage collection and distribution decisions efficiently (Carden and DelliFraine 2005; Cerveny 1980; Cohen and Pierskalla 1975; Cohen et al. 1979; Graf et al. 1972; Jacobs et al. 1996; Or and Pierskalla 1979; Price and Turcotte 1986; Sahin et al. 2007), (2) motivations to donate and factors affecting donor behaviors and donation patterns such as donor

age (Ownby et al. 1999), previous donation pattern (Flegel et al. 2000; James and Matthews 1996; Ownby et al. 1999; Yu et al. 2007), previous short-time temporary deferrals (Halperin et al. 1998; Piliavin 1987), physiological reactions (France et al. 2004), education level (Flegel et al. 2000; Schreiber et al. 2005), race and ethnicity (Schreiber et al. 2006) using statistical methods, (3) forecasting models to forecast demand and supply (Cohen et al. 1981; Cumming et al. 1976; Elston and Pickrel 1963; Frankfurter et al. 1974; Katsaliaki 2008; Pereira 2004), (4) scheduling mobile blood drives to alleviate the seasonal imbalances between the supply and demand (Cumming et al. 1976; Ghandforoush and Sen 2010; Pegels et al. 1975), and (5) the effect of extending the shelf life of whole blood and red blood cells by freezing (Cumming et al. 1977; Hess 2004; Kahn et al. 1978; Pegels et al. 1977) and using adenine (Brodheim and Hirsch 1979; Cohen et al. 1983) to balance supply and demand.

Several papers report on the benefits and successful implementation of MCA donations to improve the donor utilization, blood product availability and eliminate shortages (AuBuchon et al. 2007; Blanco 2002; Bonomo et al. 2004; Ridley 2009; Valbonesi et al. 2005; Waxman 2002), but the societal and economic benefits of MCA are neither fully explored nor quantified to develop donation tailoring protocols (Pierskalla 2004; Prastacos 1984). A few studies discuss preliminary analysis on individual products and/or individual donation types, but not for the whole system. Connelly and Pink (2002) conduct a cost-effectiveness analysis of plasma collection via two alternative methods: (1) whole blood collection and (2) erythroplasmapheresis (red blood cell-plasma) collection. Valbonesi et al. (2001) analyze collecting tailored red blood cells along with platelets during double platelet apheresis. Madden et al. (2007) investigate the impact of using a double red blood cell collection technology on the availability of the product.

Routing of blood collection vehicles in order to collect donated whole blood from blood donation centers is called *Blood Collection Problem* (BCP). BCP in general terms is first introduced by Prastacos (1984), and the objective is to collect the targeted amount of blood from donation/collection centers with minimum cost. However, they do not consider the time restriction in order to extract platelets. Blood collection operations considering the processing time limit has been studied by Doerner et al. (2008), Ghandforoush and Sen (2010) and Yi and Scheller-Wolf (2003). Ghandforoush and Sen (2010) develop a decision support system for platelet production and blood mobile scheduling, but they develop a very general model for amount collected and processed from each donation site without considering routing of the collection vehicles and the continuity of the blood donations. Assuming continuous donation at each donation site, Doerner et al. (2008) look for a minimum cost/distance solution while collecting every donated blood. Yi and Scheller-Wolf (2003) look for a minimum cost solution while collecting a pre-specified amount of blood assuming a uniform donation rate. However, as they mention, the amount of donations may change based on the time of day. For example, there may be a lunch-hour rush. In their model, each donation site is visited only once. However, as mentioned by the authors, allowing multiple visits to a donation site may increase the amount of blood collected. The challenge in the multi-visit approach is to keep track of the collected blood so that it is not counted twice (Yi and Scheller-Wolf 2003).

10.3 Future Research Opportunities

Although various aspects of blood supply chain management have been studied by several researchers, it is still a fruitful research area due to recent advances and different operational aspects. In this section, we discuss the implications of these advances and operational issues and provide research questions worth investigating.

10.3.1 Matching Supply with Demand via Donation Tailoring

Although blood has life-saving importance, excess donations are not desired due to high donation cost, inventory holding cost and disposal cost of the excess amount. The development of MCA devices allows organizations to increase donor utilization and maintain desired blood-type and -product specific inventory level of this short-supply product in a cost-effective manner by tailoring the donations based on the eligibility of donors. While determining the type of donation for each donor, one has to consider several factors such as blood-type and -product specific forecasted demand, current inventory levels, donation and inventory costs, donation pattern of individuals and deferral time between donations. If platelets have a high demand in a certain area at a certain time, donation organizations may want to encourage the donors to perform an eligible donation type that produces more platelets. Although MCA donations allow the blood donation organizations increase the donor utilization and keep a better type-inventory, how to utilize MCA donations is not fully explored to develop donation tailoring protocols. Different organizations have different strategies for donation tailoring as provided in Table 10.1.

Table 10.1 Donation tailoring strategies for different blood types from different donation organizations (SDP: single donor platelets, PLS: plasma, DPP: double platelets, RBC: red blood cells, DRBC: double red blood cells)

	Donation type				
Organization	PLS	SDP/DPP	DRBC	SDP-PLS	SDP-RBC
American Red Cross (http://www.redcrossblood.org)	AB	AB,A+,B+,O+	A−,B−,O		
Gulf Coast Regional Blood Center (http://www.giveblood.org)	AB	A	B,O		
Central Blood Bank (http://www.centralbloodbank.org)	AB	A	A,B,O		
Blood Center of Wisconsin (http://www.bcw.edu)	AB	A,B	O		
New York Blood Center (http://www.nybloodcenter.org)	AB	AB,A,B,O	B,O	AB,A	A−,B,O

From this table, we can see that there is not a clear understanding of how to utilize MCA donations except for encouraging plasma donation for AB type donors (AB is universal plasma donor) and double red blood cell donation for O type donors (O is universal red blood cell donor). While developing tailored donation schedules, there are several factors that have to be taken into account:

- *Future demand and current inventory level:* The forecasted demand and current inventory level determine the need for each blood product. The blood product(s) to be collected from donors should be determined based on the need. For example, if the platelets have a high forecasted demand and a low inventory level, platelet donations can be given higher priority, i.e., based on the eligibility of the individuals, up to triple units of platelets (Moog 2009) or 2 units of platelets and 2 units of red blood cells can be collected (Valbonesi et al. 2005).
- *Type compatibility:* Although the same type of blood is preferred for a transfusion, a different compatible type can also be used in case of an emergency. For the platelet transfusion, compatibility testing is not required. The donor-recipient compatibility for the red blood cells and plasma is provided in Tables 10.2 and 10.3. Type compatibility complicates the problem further since there is no one universal donor for all blood products.

Table 10.2 Donor-recipient compatibility for red blood cells

	Donor							
Recipient	O−	O+	A−	A+	B−	B+	AB−	AB+
O−	✓							
O+	✓	✓						
A−	✓		✓					
A+	✓	✓	✓	✓				
B−	✓				✓			
B+	✓	✓			✓	✓		
AB−	✓		✓		✓		✓	
AB+	✓	✓	✓	✓	✓	✓	✓	✓

Table 10.3 Donor-recipient compatibility for plasma

	Donor			
Recipient	O	A	B	AB
O	✓	✓	✓	✓
A		✓		✓
B			✓	✓
AB				✓

- *Shelf-life of blood products:* Since the blood products have limited shelf life, the current donations can be used to satisfy the demand for a certain amount of time.

Table 10.4 Model for improved donor utilization (SDP: single donor platelets, PLS: plasma, DPP: double platelets, DPLS: double plasma, RBC: red blood cells, DRBC: double red blood cells)

Donation type	Men(%)	Women(%)
SDP	100	95
SDP-PLS	100	95
DPP	70	51
DPP-PLS	68	22
SDP-DPLS	98	36
SDP-RBC	100	86
DPP-RBC	64	10
SDP-PLS-RBC	95	22
DRBC	94	22

- *Donation pattern of donors:* If a donor donates two times a year, then two donations with a single unit of red blood cells may be preferred over a single donation with two units of red blood cells during more stable times.
- *Donation and inventory holding cost:* Despite the live-saving importance of blood, excessive donations, and thus, perished products affect the total cost of the blood donation process by increasing the disposal, inventory and donation costs. Moreover, donating more units of one or more products increases the costs due to more staff and equipment hours.
- *Eligibility of the donors:* Not all of the donors are eligible for every donation type. However, according to results by Valbonesi et al. (2005), the acceptance rate of MCA donations is almost 100%. An example for the eligibility of donors for a specific population is presented in Table 10.4 (Keller and Bainbridge 1998).
- *Deferral times between donations:* After each donation, there is a certain amount of time that the donor has to wait before making another donation. This deferral time depends on the type of donation, and it significantly affects the subsequent donations. Deferral times are presented in Table 10.5 (Stanford Blood Center 2015; Valbonesi et al. 2005). For example, when a donor makes a DPP-PLS donation s/he has to wait for 28 days to make a SDP-DPLS donation.

We demonstrate how MCA donations can improve/change the availability of blood products by a simple example. We assume that we have 10 people donating on a given day. If these people make a whole blood donation, the total amount of products collected are 10 units of red blood cells, 5–10 units of plasma, and 1–2 units of platelets. However, using the eligibility figures in Table 10.4, we can have different scenarios based on the need. Below, we list some of the scenarios and the resulting amounts collected. Depending on the demand and collection costs, best combination of donations can be determined.

- 5 DRBC and 5 SDP-DPLS $\Rightarrow$ 10 units of RBC, 10 units of plasma, 5 units of platelets
- 5 DRBC and 5 DPP-RBC $\Rightarrow$ 15 units of RBC, 10 units of platelets
- 7 DPP and 3 DRBC $\Rightarrow$ 6 units of RBC, 14 units of platelets
- 10 SDP-DPLS $\Rightarrow$ 20 units of plasma, 10 units of platelets

Table 10.5 Deferral times between consecutive donations (in terms of days)

From-To	WB	SDP	SDP-PLS	DPP	DPP-PLS	SDP-DPLS	SDP-RBC	DPP-RBC	SDP-PLS-RBC	DRBC
WB	56	28	28	28	28	28	56	56	56	56
SDP	2	2	2	2	2	2	2	2	2	2
SDP-PLS	28	2	28	2	28	28	2	2	2	2
DPP	7	7	7	7	7	7	7	7	7	2
DPP-PLS	28	7	28	7	28	28	7	7	7	2
SDP-DPLS	56	2	56	2	56	56	2	2	2	2
SDP-RBC	56	28	28	28	28	28	56	56	56	56
DPP-RBC	56	28	28	28	28	28	56	56	56	56
SDP-PLS-RBC	56	28	28	28	28	28	56	56	56	56
DRBC	112	28	28	28	28	28	112	112	112	112

In the short-run, a very simple donation tailoring problem is as follows: Given a set of donors in a given time period (day/week), what needs to be collected from each donor to meet the demand in the next period while minimizing total collection cost? We define and formulate this version of the problem as follows. We assume that we have p donation types. Let $\mathbb{S}$ be the set of all different combinations of these p donation types except the empty set ($|\mathbb{S}| = 2^p - 1$). That is, $S \in \mathbb{S}$ means that $S \subseteq \{1, 2, \ldots, p\}$ and $S \neq \emptyset$. Let N_S be the number of donors who can do at least one of the donation types in S. Let n be the number of different blood products, and D_j be the demand of product j for $j \in \{1, 2, \ldots, n\}$. We use C_i to denote the cost of donation type i, and a_{ij} is the amount of units of product j obtained from a single occurrence of donation type i. Using decision variables x_i which is the number of type i donations performed, the problem of satisfying the demand with minimum donation cost can be formulated as follows:

$$\text{IP1: Minimize} \quad \sum_{i=1}^{p} C_i x_i \tag{10.1}$$

$$\text{subject to} \quad \sum_{i=1}^{p} a_{ij} x_i \geq D_j \qquad j \in \{1, 2, \ldots, n\} \tag{10.2}$$

$$\sum_{i \in S} x_i \leq N_S \qquad S \in \mathbb{S} \tag{10.3}$$

$$x_i \geq 0 - integer \qquad i \in \{1, 2, \ldots, p\} \tag{10.4}$$

It is not difficult to see that this is an NP-hard problem in general (reduction follows from *Integer Minimization Knapsack Problem* (Kellerer et al. 2004)). Note that in the above formulation, we have not considered the compatibility of the blood types for different blood products which makes the problem more complicated. This problem addresses the question of which donors to select to minimize the collection costs which may help donation organizations manage repeat donors. Donation organizations keep a database of the previous donors and call/invite them for a donation when the deferral time is over (Bosnes et al. 2005). This analysis will help the donation organizations determine which donors to call to schedule an appointment.

Another related problem is how to maximize the value of all donations. Given the demand and priority of different blood products, one may want to collect as much as possible to keep the inventory high. This helps the decision maker make a myopic decision which might be helpful in emergency situations such as terrorist attacks and influenza pandemic (Erickson et al. 2008; Kamp et al. 2010). In addition to the notation above, assuming w_j denotes the weight of product type j, the problem can be formulated as follows:

$$\text{IP2: Maximize} \quad \sum_{i=1}^{p} \sum_{j=1}^{n} w_j a_{ij} x_i \tag{10.5}$$

$$\text{subject to} \quad \sum_{i \in S} x_i \leq N_S \qquad S \in \mathbb{S} \tag{10.6}$$

$$x_i \geq 0 - integer \qquad i \in \{1, 2, \ldots, p\} \tag{10.7}$$

Both problems discussed above have exponential number of constraints (constraints (10.3) and (10.6)). A similar case of exponential number of constraints has been tackled for the well-known *Traveling Salesman Problem* (TSP) (Dantzig et al. 1954). Subtour elimination constraints developed by Dantzig et al. (1954) use all the subsets of the node set. Since there are exponential number of subsets, the idea proposed in the literature is to solve the problem (actually linear programming (LP) relaxation of the problem) without the subtour elimination constraints and then add them to the model step by step when they are violated. Similar to the subtour elimination constraints for TSP, these exponential number of constraints can be incorporated into the model iteratively. That is, if we remove constraints (10.3) from IP1 and constraints (10.6) from IP2, both problems can be solved relatively easily. IP1 reduces to *Integer Minimization Knapsack Problem*, and IP2 becomes a very simple problem. After eliminating the constraints, linear programming relaxation of each problem can be solved. Then, depending on the optimal solution of linear programming relaxation, constraints that are violated can be added to the model.

These are simple versions of a multi-period problem. In the multi-period version of the problem, the perishability of blood products and the deferral times between donations change the problem structure significantly. Moreover, in that case, inventory holding and disposal costs become crucial factors affecting the solution. Considering the factors listed above, the relevant research questions are as follows:

- Given a forecasted demand for the next period, what needs to be collected from each donor type? This is a short-term problem and helps the decision maker make a myopic decision in emergency situations.
- Given the donation pattern and the demand throughout the year, what should be collected from each donor type? This solves a long-term problem considering the effect in the future periods and helps donation organizations develop donation tailoring protocols. In a more stable environment, this will provide general guidelines, but in highly volatile demand cases, more customized donation schedules can be generated. To increase the number of collections, blood centers develop strategies to encourage current donors to donate more often (Glynn et al. 2001; Westphal 1997; Wu et al. 2001; Sanchez et al. 2001). In 2006, around 70% of donors were repeat donors (American Association of Blood Banks 2007). These long-term planning models will also help plan and schedule the donations from repeat donors considering the deferral times between donations.
- Due to the stochastic nature of the problem (e.g., accidents, natural disasters, etc.), forecasts may not be reliable. Hence, for both of the problems above, the case where the demand is uncertain can be another research question. In that case, stochastic and/or robust optimization techniques can be used in order handle the uncertainty.

Exact solution approaches can be developed for models IP1 and IP2 as discussed above. However, the problem in the multi-period setting is more challenging due to the deferral time between consecutive donations of a donor and the perishability of the blood products. Hence, looking for heuristic approaches in the multi-period setting might be more appropriate.

10.3.2 Scheduling Collection Operations

Platelets are used for cancer patients, transplant patients and people with blood disorders such as aplastic anemia. Although MCA devices provide an opportunity to collect platelets only, a large portion of the platelet demand is still satisfied through processing donated whole blood. According to the regulations imposed by FDA, blood donation organizations must process the whole blood within 6 h of donation time in order to extract platelets. Due to equipment required and the necessary storage conditions, the donated units are delivered to a central processing center/headquarter and processed there. For example, American Red Cross (ARC), organized into 36 blood regions nationwide (American Red Cross Biomedical Services 2015), supplies around 40% of the U.S. blood supply by collecting blood at mobile blood drives and fixed donation sites. These donated units are delivered to a processing center/headquarter in order to separate them into components. Each donation site is assigned to a single processing center/headquater, and they deliver the donated blood to this center for processing. For example, all the blood collected by ARC in Buffalo, NY is sent to Rochester, NY for processing, and then distributed to hospitals in Buffalo (American Red Cross Blood Services 2015). Similarly, all the blood collected by Gulf Coast Regional Blood Center in several fixed and mobile donation sites in Texas Gulf Coast, Brazos Valley and East Texas area are sent to the headquarters in Houston, TX for processing and distributed to hospitals (Gulf Coast Regional Blood Center 2016).

Considering the current practice described above, first we define the general *Blood Collection Problem* (BCP) and then discuss the differences between BCP and classical routing problems. Based on the regulations imposed by FDA, and the current setting of the blood donation and processing centers, BCP is defined as routing a fleet of vehicles to collect the donated blood from donation sites and deliver them to a processing center within 6 h of donation time. The donations are assumed to occur continuously during the day. BCP is a variant of the well-known *Vehicle Routing Problem* (VRP). The novel part of this problem is that blood donated at a donation site must be delivered to the processing center (in order to extract platelets) within a certain amount of time after the donation, and the blood accumulates over time which forces the donation organizations to schedule more than one pickup from a donation site.

Classical VRP is defined as the problem of serving a set of geographically dispersed customers with given demand from a central depot with a fleet of capacitated vehicles while minimizing the total transportation cost. In practice, there are several variants of this problem due to constraints and rules in real-life applications such as time windows and total tour length. VRP is first introduced by Dantzig and Ramser (1959), and then studied by several authors. In general, the solution approaches can be classified into three main categories: (1) exact algorithms (Baldacci et al. 2008; Fukasawa et al. 2006; Lysgaard et al. 2004; Ralphs et al. 2003), (2) classical heuristic algorithms (Clarke and Wright 1964; Gillett and Miller 1974; Laporte and Semet 2002; Renaud et al. 1996), and (3) metaheuristic algorithms

(Gendreau et al. 1994; Osman 1993; Pisinger and Ropke 2007). Toth and Vigo (2002a) provides a survey of the exact algorithms for VRP and compare their performances. Cordeau et al. (2002) provides a survey of the heuristic algorithms, and Golden et al. (2008), Laporte (2009) and Toth and Vigo (2002b) report recent surveys and challenges about VRP. Two main differences between VRP and BCP are the processing time limit and the accumulating behavior of the blood donations. Since the blood donations are performed during the operating hours and there is a processing time limit, one cannot wait until all the donations are completed. Continuous pickups from the donation sites have to be scheduled.

Although variants of BCP have been studied in the literature, these studies do not consider certain aspects of the problem. The main shortcoming of the study by Doerner et al. (2008) is that even a single donated unit has to be collected. However, not all of the donated units are used for platelet production (American Association of Blood Banks 2007). Moreover, they do not consider the availability of the collection vehicles. In Yi and Scheller-Wolf (2003), as discussed above, the single pickup assumption from a donation site and the uniformity of the donations are the two major concerns. Depending on the donation times, a feasible solution under the single-visit assumption may not exist. Furthermore, collecting/processing as much blood (and platelet accordingly) as possible is a more realistic objective due to importance of platelets in life-saving situations and the short shelf-life of platelets.

Considering the real-life implementation of the blood collection operations, one of the research questions is, given a fleet of collection vehicles how to design multi-visit collection routes to maximize the amount collected while keeping track of the already collected blood. Multiple visits will allow the donation organizations collect/process more blood, and will increase the availability of blood products. Removing any simplifying assumptions about the donation patterns during the day will provide more general solution approaches applicable to any donation pattern. We call this problem the *Maximum Blood Collection Problem* (MBCP).

We provide a formal definition MBCP as follows: Given a graph $G = (V_0, E)$ where $V_0 = \{0, 1, 2, \ldots, n\}$ is the set of donation sites ($V = \{1, 2, \ldots, n\}$), and the processing center (0), E is the set of edges connecting the nodes in V_0. For each edge $(i, j) \in E$ for some $i, j \in V_0$, there is an associated travel time t_{ij}. Associated with each donation site i for $i \in V$, there are opening (a_i) and closing (b_i) times. The donations for donation site i are assumed to occur at discrete times between a_i and b_i. We use d_t^i to denote the amount of blood donated at time $t(\in (a_i, b_i])$ at donation site i. Donated blood has to be delivered to processing center by a collection vehicle within K time units of donation time. Assuming we have M uncapacitated vehicles, the objective is to collect/process maximum amount of blood before spoilage.

In Fig. 10.2, we provide an example to illustrate MBCP and how differences between MBCP and other classical routing problems affect the solution. In this example, we have a single vehicle located at the depot/processing center (node 0), and there are two donation centers. It takes 1 h to travel from one node to another. Each donation center has a donation rate of 1 unit per hour. We assume that donations occur between 8am and 2pm, and the donated units have to be sent

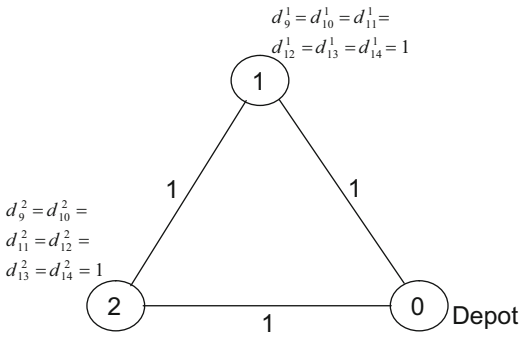

Fig. 10.2 An example illustrating the difference between MBCP and other routing problems

to processing center within 5 h of donation (allowing 1 h for processing) in order to process it before spoilage. At the end of each hour, donation of one unit of blood is completed, and this single unit becomes available for pickup. The total amount collected in each site is 6 units. If the vehicle is dispatched at 8am by assuming all the donations are ready to be collected at 9am, then we collect only 3 units in the first tour. Even if the vehicle is dispatched at 1pm in order to wait for all the units to be collected, we can collect only 5 units of blood due to spoilage time restriction. Moreover, the maximum amount of blood that can be collected in a single tour is 9 units which can be collected if the vehicle is dispatched at 10am or 11am or 12pm. Furthermore, if the objective is to collect maximum amount, then the vehicle has to visit the collection sites more than once. For example, if the vehicle is dispatched at 10am for the first tour and at 1pm for the second tour, then all the units can be collected before spoilage.

In the general definition of MBCP provided above, a donation site can be visited by more than one collection vehicle. However, due to practicality issues, a solution where each donation site is visited by the same vehicle/driver might be preferred. This results in a clustered setting of the donation sites where the set of donation sites is partitioned into clusters, each vehicle is assigned to a cluster of donation sites and visits the donation sites in that cluster only. We provide a mathematical model for the clustered setting of MBCP. First, we define additional notation needed in the formulation. The start time (T_1) and end time (T_2) of the planning horizon are defined as follows: $T_1 := \min_{i \in V} a_i$, $T_2 := \max_{i \in V} b_i$. Furthermore, we need an upper bound on the number of tours that can performed by a vehicle in a day. We use the following upper bound in the model:

$$L = \frac{T_2 - T_1 + 2K}{\min_{i \in V}\{2t_{0i}\}} \tag{10.8}$$

This upper bound uses the minimum distance from the processing center to a donation site to find a limit on the number of tours. Finally, $\mathscr{T}$ is used to denote the set of tours performed by a vehicle and defined as follows: $\mathscr{T} := \{1, 2, \ldots, L\}$.

The decision variables used in the model are as follows:

$$x_{mk}^{it} = \begin{cases} 1, \text{ if donations completed at time } t \text{ at donation} \\ \quad \text{site } i \text{ are picked up in tour } m \text{ of vehicle } k \\ 0, \text{ otherwise} \end{cases} \qquad \begin{array}{r} \forall i \in V, t \in \{a_i, a_i + 1, \ldots, b_i\}, \\ m \in \mathscr{T}, k \in \{1, 2, \ldots, M\} \end{array}$$

$$y_{ijmk} = \begin{cases} 1, \text{ if location } j \text{ is visited after} \\ \quad \text{location } i \text{ in tour } m \text{ of vehicle } k \\ 0, \text{ otherwise} \end{cases} \qquad \forall i, j \in V_0, m \in \mathscr{T}, k \in \{1, 2, \ldots, M\}$$

$$z_{mk} = \begin{cases} 1, \text{ if vehicle } k \text{ performs tour } m \\ 0, \text{ otherwise} \end{cases} \qquad \forall m \in \mathscr{T}, k \in \{1, 2, \ldots, M\}$$

e_{mk} = Arrival time of vehicle k at the processing center after completing tour m
$\forall m \in \mathscr{T}, k \in \{1, 2, \ldots, M\}$

v_{imk} = Arrival time of vehicle k at donation site i in tour m
$\forall i \in V, m \in \mathscr{T}, k \in \{1, 2, \ldots, M\}$

In the proposed mathematical model, the visiting time of a donation site (v_{imk}) in a given tour is set to the visiting time of the last visit to this site if the donation site is not visited in this tour, and H is used as a large enough number. In the following mathematical model, we maximize the total number of donations collected/processed:

$$\text{MIP: Maximize} \quad \sum_{m \in \mathscr{T}} \sum_{k \in \{1,2,\ldots,M\}} \sum_{i \in V} \sum_{t \in \{a_i, \ldots, b_i\}} d_t^i x_{mk}^{it} \tag{10.9}$$

$$\text{subject to} \quad \sum_{j \in V_0 \setminus \{i\}} y_{ijmk} = \sum_{j \in V_0 \setminus \{i\}} y_{jimk} \qquad \forall i, m, k \tag{10.10}$$

$$\sum_{j \in V_0 \setminus \{i\}} y_{ijmk} \leq 1 \qquad \forall i, m, k \tag{10.11}$$

$$\sum_{j \in V} y_{0jmk} + \sum_{j \in V} y_{j0mk} = 2z_{mk} \qquad \forall m, k \tag{10.12}$$

$$z_{mk} \leq z_{m-1,k} \qquad \forall m, k \tag{10.13}$$

$$\sum_{i \in V_0} \sum_{j \in V_0} y_{ijmk} \leq H z_{mk} \qquad \forall m, k \tag{10.14}$$

$$\sum_{j \in V_0} (y_{jimk} + \sum_{n \in \mathscr{T}} \sum_{l \in \{1,2,\ldots,M\}, l \neq k} y_{jinl}) \leq 1 \qquad \forall i, m, k \tag{10.15}$$

$$v_{imk} + t_{ij} \leq v_{jmk} + H(1 - y_{ijmk}) \qquad \forall i, j, m, k \tag{10.16}$$

$$e_{m-1,k} + t_{0i} \leq v_{imk} + H(1 - y_{0imk}) \qquad \forall i, m, k \tag{10.17}$$

$$v_{i1k} \geq a_i \qquad \forall i, k \tag{10.18}$$

$$a_i + H \sum_{j \in V_0} y_{ji1k} \geq v_{i1k} \qquad \forall i, k \qquad (10.19)$$

$$v_{i,m-1,k} + H \sum_{j \in V_0} y_{jimk} \geq v_{imk} \qquad \forall i, m, k \qquad (10.20)$$

$$v_{imk} \geq v_{i,m-1,k} \qquad \forall i, m, k \qquad (10.21)$$

$$v_{imk} + t_{i0} \leq e_{mk} + H(1 - y_{i0mk}) \qquad \forall i, m, k \qquad (10.22)$$

$$x^{it}_{mk} \leq \sum_{j \in V_0} y_{jimk} \qquad \forall i, m, k, t \qquad (10.23)$$

$$\sum_{m \in \mathscr{T}} \sum_{k \in \{1,2,\ldots,M\}} x^{it}_{mk} \leq 1 \qquad \forall i, t \qquad (10.24)$$

$$e_{mk} - K \leq t + H(1 - x^{it}_{mk}) \qquad \forall i, m, k, t \qquad (10.25)$$

$$v_{imk} \geq t - H(1 - x^{it}_{mk}) \qquad \forall i, m, k, t \qquad (10.26)$$

$$x^{it}_{mk} \in \{0, 1\} \qquad \forall i, t, m, k \qquad (10.27)$$

$$y_{ijmk} \in \{0, 1\} \qquad \forall i, j, m, k \qquad (10.28)$$

$$z_{mk} \in \{0, 1\} \qquad \forall m, k \qquad (10.29)$$

$$e_{mk} \geq 0 \qquad \forall m, k \qquad (10.30)$$

$$v_{imk} \geq 0 \qquad \forall i, m, k \qquad (10.31)$$

In MIP, objective function (10.9) maximizes the total number of donations collected and processed. Constraints (10.10)–(10.11) make sure that each donation site is visited at most once in a given tour of a vehicle. Constraints (10.12) guarantee that each vehicle returns to processing center at the end of each tour. A tour of a vehicle can only be performed if all the tours of the same vehicle before this tour are performed. This is forced by constraints (10.13). Constraints (10.14) ensure that a vehicle can visit donation sites in a tour only if such a tour is performed by the vehicle. Donation sites are clustered using constraints (10.15). Constraints (10.16)–(10.22) determine the visiting time of each donation site in each tour of each vehicle. A donation performed at a donation site can be collected at most once, and it can be collected by a vehicle visiting the donation site. This is enforced by constraints (10.23)–(10.24). Constraints (10.25)–(10.26) impose the visiting time of a donation site in a tour and the completion time of this tour if a donation at a given time is picked up from this donation site in this tour. Finally, constraints (10.27)–(10.31) impose the sign restrictions.

Due to complexity of the routing problems in general, using such a mathematical model to solve this problem has a low probability of success. Hence, heuristic algorithms are more promising for finding a good (not necessarily optimal) solution for real-life instances.

Another potential research question related to collection operations is how to design collection routes (allowing multiple visits) with minimum total routing cost while collecting/processing a certain amount of blood for platelet production. This will also help reduce soaring health care costs.

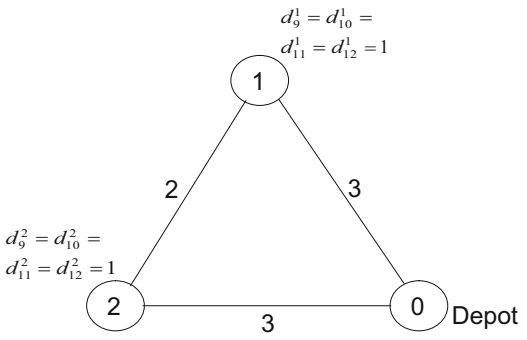

Fig. 10.3 An example illustrating the importance of synchronizing the appointment and collection schedules

Finally, in order to utilize their resources, including staff and equipment, effectively and avoid long waiting times, donation sites operate on an appointment-based schedule (American Red Cross 2011; Blood Bank of Alaska 2016; Gulf Coast Regional Blood Center 2016; LifeSource 2015). Long waiting times may discourage the donors from future donations (Brennan et al. 1992; Michaels et al. 1993). This is important due to not only the limited donor pool, but also the fact that most of the donations are from repeat donors (Brennan et al. 1992). Although the main motivation behind appointment scheduling is increasing the utilization of resources and avoiding long waiting times, appointment schedule also significantly affects the collection schedules/routes from donation sites. If the appointments are not scheduled considering the pickup schedules or the availability of collection vehicles, the donations may not be collected/processed or may increase the collection costs significantly. Moreover, due to availability of the collection vehicles, it may not be possible to collect/process every donated blood which affects the availability of blood products.

We provide a simple example (see Fig. 10.3) to illustrate how appointment and collection schedules may affect each other. In Fig. 10.3, we have a single depot/processing center, two donation sites, and a single vehicle to collect donated blood from these donation sites. The corresponding travel times (in hours) are given on each edge. We assume that each donation site is open for 4 h during 8am–12pm, and each appointment is scheduled to one of the following 1 h intervals: 8am–9am, 9am–10am, 10am–11am, 11am–12pm. Furthermore, we assume that 4 donations will occur in each donation site. If we spread the donations over time in order to have a better utilization and balanced workload, we schedule only one donation to each 1 h interval. In this case, only one donation occurs in an hour at each donation site. Finally, we assume that donated blood has to be processed within 6 h of donation time. This means that blood donated during time interval 8am–9am has to be delivered to processing center no later than 3pm. One can easily see that one vehicle cannot collect and deliver all the donated blood. The maximum amount that can be collected/processed is 6 units. To collect 6 units of blood, the vehicle has to leave the depot at 7am, and visit donation sites 1 and 2 at 10am and 12pm, respectively. However, if the collection and appointment schedules are considered

together, we can collect all the donated blood by scheduling 2 appointments to each of 8am–9am and 9am–10am intervals at donation site 1 and keeping the schedule for the second donation site same.

An integrated appointment and collection scheduling helps decrease the collection costs and/or increase the amount collected/processed by synchronizing the appointment and collection times. The research question here is as follows: What is the best time to assign a set of appointment requests considering the availability of the collection vehicles and/or the resulting collection cost? Although most of the organizations operate on an appointment-based schedule, there is still some uncertainty in the donation amounts/times due to eligibility of donors, no-shows and walk-in donors. Incorporating such uncertainties into proposed solution approaches is another interesting research direction.

10.4 Conclusion

Most of the previous research on the efficient management of blood donation operations focuses on inventory models for perishable products, the regionalization and organization of the blood banking supply chain and the statistical analysis of factors affecting wastage and repeated blood donation. Recent advances such as multicomponent apheresis (MCA) provide several opportunities including (1) increasing the donor utilization and (2) tailoring the donations based on the demand. We discuss the potential benefits of MCA and pose research questions in order to improve blood products availability and manage donation/disposal costs using MCA. Due to regulations, donated whole blood has to be processed within 6 h of donation in order to extract platelets. We elaborate on the effect of this processing time limit on the collection operations, discuss practical issues and potential research questions. We believe that the discussion in this chapter will lead to interesting research areas worth future investigation from both practical and theoretical perspectives. Moreover, considering the life-saving importance of blood, future research in this field is expected to have a significant societal impact.

Acknowledgements This research is supported in part by TUBITAK grant 112M945.

References

Abbott RD, Friedman BA, Williams GW (1978) Recycling older blood by integration into the inventory of a single large hospital blood bank: a computer simulation approach. Transfusion 18(6):709–715

Albright SC (1976) Optimal stock depletion policies with stochastic lives. Manag Sci 22(8):852–857

Alshamrani A, Mathur K, Ballou RH (2007) Reverse logistics: simultaneous design of delivery routes and returns strategies. Comput Oper Res 34(2):595–619

American Association of Blood Banks (2007) National blood collection and utilization survey
American Association of Blood Banks (2011) National blood collection and utilization survey
American Association of Blood Banks (2015) http://www.aabb.org
American Red Cross (2011) http://www.redcrossblood.org
American Red Cross Biomedical Services (2015) A leader in providing lifesaving blood and blood products to the nation. http://www.redcrossblood.org
American Red Cross Blood Services (2015) http://volunteer.united-e-way.org/uwbec/org/567730.html
AuBuchon JP, Dumont LJ, Herschel L, Roger J, Beddard RL, Taylor HL, Whitley PH, Sawyer SL, Graminske S, Martinson K, Dora R, Heldke S, Adamson J, Rose LE (2007) Automated collection of double red blood cell units with a variable-volume separation chamber. Transfusion 48 (1):147–152
Baldacci R, Christofides N, Mingozzi A (2008) An exact algorithm for the vehicle routing problem based on the set partitioning formulation with additional cuts. Math Program 115(2):351–385
Belien J, Force, H (2012) Supply chain management of blood products: a literature review. Eur J Oper Res 217(1):1–16
Blanco L (2002) Tailored collection of multicomponent by apheresis. Transfus Apher Sci 27(2): 123–127
Blood Bank of Alaska (2016) http://www.bloodbankofalaska.org
Bonomo P, Garozzo G, Bennardello F (2004) The selection of donors in multicomponent collection management. Transfus Apher Sci 30(1):55–59
Bosnes V, Aldrin M, Heier HE (2005) Predicting blood donor arrival. Transfusion 45(2):162–170
Brennan JE, Golden BL, Rappoport HK (1992) Go with the flow: improving Red Cross bloodmobiles using simulation analysis. Interfaces 22(5):1–13
Brodheim E, Hirsch R (1979) Effect of adenine on blood usage strategies, shortages and outdating. Transfusion 19(2):105–106
Brodheim E, Prastacos GP (1979a) The Long Island blood distribution system as a prototype for regional blood management. Interfaces 9(5):3–20
Brodheim E, Prastacos GP (1979b) A regional blood management system with prescheduled deliveries. Transfusion 19(4):455–462
Brodheim E, Derman C, Prastacos G (1975) On the evaluation of a class of inventory policies for perishable products such as blood. Manag Sci 21(11):1320–1325
Brodheim E, Hirsch F, Prastacos G (1976) Setting inventory levels for hospital blood banks. Transfusion 16(1):63–70
Carden R, DelliFraine JL (2005) An examination of blood center structure and hospital customer satisfaction. Health Mark Q 22(3):21–42
Cerveny RP (1980) An application of warehouse location techniques to bloodmobile operations. Interfaces 10(6):88–96
Chazan D, Gal S (1977) A Markovian model for a perishable product inventory. Manag Sci 23(5): 512–521
Clarke G, Wright JW (1964) Scheduling of vehicles from a central depot to a number of delivery points. Oper Res 12(4):568–581
Cohen MA, Pierskalla WP (1975) Management policies for a regional blood bank. Transfusion 15 (1):58–67
Cohen MA, Pierskalla WP (1979) Target inventory levels for a hospital blood bank or a decentralized regional blood banking system. Transfusion 19(4):444–454
Cohen MA, Pierskalla WP, Sassetti RJ, Consolo J (1979) An overview of a hierarchy of planning models for regional blood bank management. Transfusion 19(5):526–534
Cohen MA, Pierskalla WP, Sassetti RJ (1981) Regionalization of blood banking services. Technical report, University of Pennsylvania
Cohen MA, Pierskalla WP, Sassetti RJ (1983) The impact of adenine and inventory utilization decisions on blood inventory management. Transfusion 23(1):54–58
Connelly LB, Pink A (2002) An economic evaluation of plasma production via erythroplasmapheresis and whole blood collection. Transfus Apher Sci 27(2):101–111

Cordeau JF, Gendreau M, Laporte G, Potvin JY, Semet F (2002) A guide to vehicle routing heuristics. J Oper Res Soc 53(5):512–522

Cumming PD, Kendall KE, Pegels CC, Seagle JP, Shubsda JF (1976) A collections planning model for regional blood suppliers: description and validation. Manag Sci 22(9):962–971

Cumming PD, Kendall KE, Pegels CC, Seagle JP (1977) Cost effectiveness and use of frozen blood to alleviate blood shortages. Transfusion 17(6):602–606

Currie CJ, Patel TC, McEwan P, Dixon S (2004) Evaluation of the future supply and demand for blood products in the United Kingdom National Health Service. Transfus Med 14(1):19–24

Dantzig GB, Ramser JH (1959) The truck dispatching problem. Manag Sci 6(1):80–91

Dantzig G, Fulkerson D, Johnson S (1954) Solution of a large-scale traveling-salesman problem. Oper Res 2:393–410

Davey RJ (2004) Recruiting blood donors: challenges and opportunities. Transfusion 44(4):597–600

Denesiuk L, Richardson T, Nahirniak S, Clarke G (2006) Implementation of a redistribution system for near-outdate red blood cell units. Arch Pathol Lab Med 130(8):1178–1183

Doerner KF, Gronalt M, Hartl RF, Kiechle G, Reimann M (2008) Exact and heuristic algorithms for the vehicle routing problem with multiple interdependent time windows. Comput Oper Res 35(9):3034–3048

Dumas MB, Rabinowitz M (1977) Policies for reducing blood wastage in hospital blood banks. Manag Sci 23(10):1124–1132

Elston RC, Pickrel JC (1963) A statistical approach to ordering and usage policies for a hospital blood bank. Transfusion 3(1):41–47

Elston RC, Pickrel JC (1965) Guides to inventory levels for a hospital blood bank determined by electronic computer simulation. Transfusion 5(5):465–470

Erickson ML, Champion MH, Klein R, Ross RL, Neal ZM, Snyder EL (2008) Management of blood shortages in a tertiary care academic medical center: the Yale-New Haven hospital frozen blood reserve. Transfusion 48(10):2252–2263

Federgruen A, Prastacos G, Zipkin PH (1986) An allocation and distribution model for perishable products. Oper Res 34(1):75–82

Flegel WA, Besenfelder W, Wagner FF (2000) Predicting a donor's likelihood of donating within a preselected time interval. Transfus Med 10(3):181–192

Fontaine MJ, Chung YT, Rogers WM, Sussmann HD, Quach P, Galel SA, Goodnough LT, Erhun F (2009) Improving platelet supply chains through collaborations between blood centers and transfusion services. Transfusion 49(10):2040–2047

France CR, France JL, Roussos M, Ditto B (2004) Mild reactions to blood donation predict a decreased likelihood of donor return. Transfus Apher Sci 30(1):17–22

Frankfurter GM, Kendall KE, Pegels CC (1974) Management control of blood through a short-term supply-demand forecast system. Manag Sci 21(4):444–452

Friedman BA, Abbott RD, Williams GW (1982) A blood ordering strategy for hospital blood banks derived from a computer simulation. Am J Clin Pathol 78(2):154–160

Fries BE (1975) Optimal ordering policy for a perishable commodity with fixed lifetime. Oper Res, 23(1):46–61

Fukasawa R, Longo H, Lysgaard J, de Aragao MP, Reis M, Uchoa E, Werneck FR (2006) Robust branch-and-cut-and-price for the capacitated vehicle routing problem. Math Program 106(3): 491–451

Gendreau M, Hertz A, Laporte G (1994) A tabu search heuristic for the vehicle routing problem. Manag Sci 40(10):1276–1290

Ghandforoush P, Sen TK (2010) A DSS to manage platelet production supply chain for regional blood centers. Decis Support Syst 50(1):32–42

Gilcher RO, McCombs S (2005) Seasonal blood shortages can be eliminated. Curr Opin Hematol 12(6):503–508

Gillett BE, Miller LR (1974) A heuristic algorithm for the vehicle dispatch problem. Oper Res 22 (2):340–349

Glynn SA, Smith JW, Schreiber GB, Kleinman SH, Nass, CC, Bethel J, Biswas B, Thomson RA, Williams AE (2001) Repeat whole blood and plateletpheresis donors: unreported deferrable risks, reactive screening tests, and response to incentive programs. Transfusion 41(6):736–743
Goh C-H, Greenberg BS, Matsuo H (1993) Two-stage perishable inventory models. Manag Sci 39 (5):633–649
Golden BL, Raghavan S, Wasil EA (2008) The vehicle routing problem: latest advances and new challenges. Operations research computer science interfaces series, vol 43. Springer, New York
Goyal SK, Giri BC (2001) Recent trends in modeling of deteriorating inventory. Eur J Oper Res 134(1):1–16
Graf S, Katz A, Morse E (1972) Blood inventory distribution by type in a total supply system. Transfusion 12(3):185–189
Gregor PJ, Forthofer RN, Kapadia AS (1982) An evaluation of inventory and transportation policies of a regional blood distribution system. Eur J Oper Res 10(1):106–113
Gulf Coast Regional Blood Center (2016) http://www.giveblood.org
Haemonetics Corporation (2008) Transforming blood management. 2008 Annual Report
Haijema R, Van Der Wal J, Van Dijk NM (2007) Blood platelet production: optimization by dynamic programming and simulation. Comput Oper Res 34(3):760–779
Haijema R, Van Dijk N, Van Der Wal J, Sibinga, CS (2009) Blood platelet production with breaks: optimization by SDP and simulation. Int J Prod Econ 121(2):464–473
Halperin D, Baetens J, Newman B (1998) The effect of short-term, temporary deferral on future blood donation. Transfusion 38(2):181–183
Hemmelmayr V, Doerner KF, Hartl RF, Savelsbergh MWP (2009) Delivery strategies for blood products supplies. OR Spectr 31(4):707–725
Hemmelmayr V, Doerner KF, Hartl RF, Savelsbergh MWP (2010) Vendor managed inventory for environments with stochastic product usage. Eur J Oper Res 202(3):686–695
Hess JR (2004) Red cell freezing and its impact on the supply chain. Transfus Med 14(1):1–8
Hesse SM, Coullard CR, Daskin MS, Hurter AP (1997) A case study in platelet inventory management. In: Proceedings of the Sixth Annual Industrial Engineering Research Conference, pp 801–806
Hurlburt EL, Jones AR (1964) Blood bank inventory control. Transfusion 4(2):126–133
Jacobs DA, Sylan MN, Clemson BA (1996) An analysis of alternative locations and service areas of American Red Cross blood facilities. Interfaces 26(3):40–50
Jagannathan R, Sen T (1991) Storing crossmatched blood: a perishable inventory model with prior allocation. Manag Sci 37(3):251–266
James RC, Matthews DE (1996) Analysis of blood donor return behaviour using survival regression methods. Transfus Med 6(1):21–30
Jennings JB (1968) An analysis of hospital blood bank whole blood inventory control policies. Transfusion 8(6):335–342
Jennings JB (1973) Blood bank inventory control. Manag Sci 19(6):637–645
Kahn RA, McDonough B, Rowe A, Ellis FR, Pino B (1978) The impact of converting to an all frozen blood system in a large regional blood center. Transfusion 18(3):304–311
Kamp C, Heiden M, Henseler O, Seitz R (2010) Management of blood supplies during an influenza pandemic. Transfusion 50(1):231–239
Karaesmen IZ, Scheller-Wolf A, Deniz B (2011) Managing perishable and aging inventories: review and future research directions. In: Planning production and inventories in the extended enterprise. International series in operations research and management science. Kluwer Academic, Dordrecht, pp 393–436
Kaspi H, Perry D (1983) Inventory systems of perishable commodities. Adv Appl Probab 15(3): 674–685
Katsaliaki K (2008) Cost-effective practices in the blood service sector. Health Policy 86(2):276–287
Katz AJ, Carter CW, Saxton P, Blutt J, Kakaya RM (1983) Simulation analysis of platelets production and inventory management. Vox Sang 44(1):31–36

Keller NC, Bainbridge MA (1998) Model for improved donor utilization by multiple product collection. Transfusion 38:S26
Kellerer H, Pferschy U, Pisinger D (2004) Knapsack problems. Springer, Berlin
Kendall KE (1980) Multiple objective planning for regional blood centers. Long Range Plan 13 (4):98–104
Kendall KE, Lee SM (1980) Formulating blood rotation policies with multiple objectives. Manag Sci 26(11):1145–1157
Kopach R, Balcioglu B, Carter M (2008) Tutorial on constructing a red blood cell inventory management system with two demand rates. Eur J Oper Res 185(3):1051–1059
Laporte G (2009) Fifty years of vehicle routing. Transp Sci 43(4):408–416
Laporte G, Semet F (2002) Classical heuristics for the capacitated VRP. In The vehicle routing problem. SIAM monographs on discrete mathematics and applications, Society for Industrial and Applied Mathematics, Philadelphia, pp 109–128
Ledman RE, Groh N (1984) Platelet production planning to ensure availability while minimizing outdating. Transfusion 24(6):532–533
LifeSource (2015) http://www.lifesource.org
Lysgaard J, Letchford AN, Eglese RW (2004) A new branch-and-cut algorithm for the capacitated vehicle routing problem. Math Program 100(2):423–445
Madden E, Murphy EL, Custer B (2007) Modeling red cell procurement with both double-red-cell and whole-blood collection and the impact of European travel deferral on units available for transfusion. Transfusion 47(11):2025–2037
Matthes GA (2002) Options and cost effectiveness of multicomponent blood collection. Transfus Sci 27(2):115–121
Mendez A, Wagli F, Schmid I, Frey BM (2007) Frequent platelet apheresis donations in volunteer donors with hemoglobin < 125g/l are safe and efficient. Transfus Apher Sci 36(1):47–53
Michaels JD, Brennan JE, Golden BL, Fu MC (1993) A simulation study of donor scheduling system for the American Red Cross. Comput Oper Res 20(2):199–213
Moog R (2009) Feasibility and safety of triple dose platelet collection by apheresis. J Clin Apher 24(6):238–240
Nahmias S (1975) Optimal ordering policies for perishable inventory-II. Oper Res 23(4):735–749
Nahmias S (1976) Myopic approximations for the perishable inventory problem. Manag Sci 22(9): 1002–1008
Nahmias S (1982) Perishable inventory theory: a review. Oper Res 30(4):680–708
Nahmias S, Pierskalla WP (1973) Optimal ordering policies for product that perishes in two periods subject to stochastic demand. Nav Res Logist 20(2):207–229
Or I, Pierskalla WP (1979) A transportation location-allocation model for regional blood banking. AIIE Trans 11(2):86–95
Osman IH (1993) Metastrategy simulated annealing and tabu search algorithms for the vehicle routing problem. Ann Oper Res 41(4):421–451
Ownby HE, Kong F, Watanabe K, Tu Y, Nass CC (1999) Analysis of donor return behavior. Transfusion 39(10):1128–1135
Pegels CC, Jelmert AE (1970) An evaluation of blood-inventory policies: a Markov chain application. Oper Res 18(6):1087–1098
Pegels CC, Jelmert AE (1971) A study of two blood bank crossmatch policies. AIIE Trans 3(1): 69–75
Pegels CC, Seagle JP, Cumming PD, Kendall KE (1975) A computer-based interactive planning system for scheduling blood collections. Transfusion 15(4):381–386
Pegels CC, Seagle JP, Cumming PC, Kendall KE, Shubsda JF (1977) An analysis of selected blood service policy changes. Med Care 15(2):147–157
Pereira A (2004) Performance of time-series methods in forecasting the demand for red blood cell transfusion. Transfusion 44(5):739–746
Pereira A (2005) Blood inventory management in the type and screen era. Vox Sang 89(4):245–250
Perry D, Posner MJM (1990) Control of input and demand rates in inventory systems of perishable commodities. Nav Res Logist 37(1):85–97

Pierskalla WP (2004) Supply chain management of blood banks. In: Operations research and health care: a handbook of methods and applications. Kluwer Academic, New York, pp 103–145
Pierskalla WP, Roach CD (1972) Optimal issuing policies for perishable inventory. Manag Sci 18 (11):603–614
Piliavin JA (1987) Temporary deferral and donor return. Transfusion 27(2):199–200
Pisinger D, Ropke S (2007) A general heuristic for vehicle routing problems. Comput Oper Res 34(8):2403–2435
Pitocco C, Sexton TR (2005) Alleviating blood shortages in a resource-constrained environment. Transfusion 45(7):1118–1126
Popovsky MA (2005) Multicomponent apheresis blood collection in the United States: current status and future directions. Transfus Apher Sci 32(3):299–304
Prastacos GP (1978) Optimal myopic allocation of a product with fixed lifetime. J Oper Res Soc 29(9):905–913
Prastacos GP (1981) Allocation of a perishable product inventory. Oper Res 29(1):95–107
Prastacos GP (1984) Blood inventory management: an overview of theory and practice. Manag Sci 30(7):777–780
Prastacos GP, Brodheim E (1979) Computer-based regional blood distribution. Comput Oper Res 6(2):69–77
Prastacos GP, Brodheim E (1980) PBDS: a decision support system for regional blood management. Manag Sci 26(5):451–463
Price WL, Turcotte M (1986) Locating a blood bank. Interfaces 16(5):17–26
Rabinowitz M (1973) Blood bank inventory policies: a computer simulation. Health Serv Res 8 (4):271–282
Ralphs T, Kopman L, Pulleyblank W, Trotter L (2003) On the capacitated vehicle routing problem. Math Program 94(2–3):343–359
Renaud J, Boctor FF, Laporte G (1996) An improved petal heuristic for the vehicle routing problem. J Oper Res Soc 47(2):329–336
Ridley J (2009) Improving the management of the blood supply through the use of Haemonetics' MCS-8150 automated double red cell collection devices. Transfus Apher Sci 41(1):39–43
Rytila JS, Spens KM (2006) Using simulation to increase efficiency in blood supply chains. Manag Res News 29(12):801–819
Sahin G, Sural H, Meral S (2007) Locational analysis for regionalization of Turkish Red Crescent blood services. Comput Oper Res 34(3):692–704
Sanchez AM, Ameti DI, Schreiber GB, Thomson RA, Lo A, Bethel J, Williams AE (2001) The potential impact of incentives on future blood donation behavior. Transfusion 41(2):172–178
Sapountzis C (1984) Allocating blood to hospitals from a central blood bank. Eur J Oper Res 16 (2):157–162
Sapountzis C (1989) Allocating blood to hospitals. J Oper Res Soc 40(5):443–449
Schreiber GB, Sharma UK, Wright DJ, Glynn SA, Ownby HE, Tu Y, Garratty G, Piliavin J, Zuck T, Gilcher R (2005) First year donation patterns predict long-term commitment for first-time donors. Vox Sang 88(2):114–121
Schreiber GB, Schlumpf KS, Glynn SA, Wright DJ, Tu Y, King MR, Higgins MJ, Kessler D, Gilcher R, Nass CC, Guiltinan AM (2006) Convenience, the bane of our existence, and other barriers to donating. Transfusion 46(4):545–553
Sirelson V, Brodheim E (1991) A computer planning model for blood platelet production and distribution. Comput Methods Programs Biomed 35(4):279–291
Smith JW, Gilcher RO (2006) The future of automated red blood cell collection. Transfus Apher Sci 24(2):219–226
South Texas Blood & Tissue Center (2014) http://www.southtexasblood.org
Stanford Blood Center (2015) http://bloodcenter.stanford.edu
Toth P, Vigo D (2002a) Models, relaxations and exact approaches for the capacitated vehicle routing problem. Discret Appl Math 123(1–3):487–512
Toth P, Vigo D (eds) (2002b) The vehicle routing problem. SIAM monographs on discrete mathematics and applications. Society for Industrial and Applied Mathematics, Philadelphia

Valbonesi M, Bruni R, Bo A, De Luigi MC, Stura P, Carlier P, Sanfilippo B (2001) Double plateletpheresis (DPA) and tailored RBC collection with the Excel-Pro: preliminary results. Transfus Apher Sci 24(1):71–73

Valbonesi M, Giannini G, Morelli F, Frisoni R, Capra C (2005) Multicomponent collection as of 2005. Transfus Apher Sci 32(3):287–297

Van Dijk N, Haijema R, Van Der Wal J, Sibinga CS (2009) Blood platelet production: a novel approach for practical optimization. Transfusion 49(3):411–420

Vaquero MAP (2006) Multicomponent collection: the basque experience. Transfus Apher Sci 34 (1):84–86

Vrat P, Khan AB (1976) Simulation of a blood-inventory-bank system in a hospital. Socio Econ Plan Sci 10(1):7–15

Waxman DA (2002) Volunteer donor apheresis. Ther Apher 6(1):77–81

Westphal RG (1997) Donors and the United States blood supply. Transfusion 37(2):237–241

Williams AE, Thomson RA, Schreiber GB, Watanabe K, Bethel J, Lo A, Kleinman SH, Hollingsworth CG, Nemo GJ (1997) Estimates of infectious disease risk factors in US blood donors. Retrovirus epidemiology donor study. J Am Med Assoc 277(12):967–972

World Health Organization (2016) http://www.who.int

Wu Y, Glynn SA, Schreiber GB, Wright DJ, Lo A, Murphy EL, Kleinman SH, Garratty G (2001) First-time blood donors: demographic trends. Transfusion 41(5):360–364

Yahnke DP, Rimm AA, Mundt CJ, Aster R, Hurst TM (1972) Analysis and optimization of a regional blood bank distribution process. Transfusion 12(2):111–118

Yen H (1975) Inventory management for a perishable product multi-echelon system. PhD thesis, Northwestern University

Yi J, Scheller-Wolf A (2003) Vehicle routing with time windows and time-dependent rewards: a problem from the American Red Cross. Technical report, Carnegie Mellon University

Yu PL, Chung KH, Lin CK, Chan JS, Lee CK (2007) Predicting potential drop-out and future commitment for first-time donors based on first 1.5-year donation patterns: the case in Hong Kong Chinese donors. Vox Sang 93(1):57–63

Chapter 11
Vaccine Supply Management

Nafiseh Shamsi G. and S. Ali Torabi

11.1 Introduction

There are several drivers in the world such as natural or man-made disasters, malnutrition, population growth and environmental change for triggering different infectious diseases which increase the risk of morbidity and mortality. Delay in control of such problems may cause epidemic outbreaks. The main objectives of health organizations to containment epidemics include reducing the transmission rate, reducing the mortality rate of infected individuals and increasing the immunization rate of susceptible individuals by vaccination or quarantine programs as control tools in addition to other public health interventions such as providing clean water, better nutrition and medical supplies.

Control tools could be adopted with the aim of either preventing the spread after the initiation of an infectious disease (i.e. pre-event action) for which a certain level of medical supplies should be kept in order to take immediate action, or to control a confirmed epidemic (i.e. post-event action) (Dasaklis et al. 2012).

Therefore, timely control of an epidemic outbreak strongly depends on the establishment of an emergency supply chain adopted at international, national or community level.

On the other hand, introducing new vaccines that target such diseases as typhoid, malaria, and dengue and their availability in low or middle-income countries need a supply chain view (Kaufmann et al. 2011) which is currently facing serious challenges in developing countries (Zaffran et al. 2013). Therefore, designing appropriate vaccine supply chain (VSC) networks has a significant role to overcome these challenges.

N. Shamsi G. • S.A. Torabi (✉)
School of Industrial Engineering, University of Tehran, Tehran, Iran
e-mail: satorabi@ut.ac.ir

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_11

Internationally, the United Nations International Children's Emergency Fund (UNICEF) supplies vaccines reaching 40% of the world's children as part of its commitment to improving child survival. The procurement of vaccines and related supplies is UNICEF's largest procurement activity. In 2014, UNICEF procured 2.71 billion doses of vaccines for 100 countries, from polio and measles to tetanus, Bacillus Calmette–Guérin (BCG) and yellow fever vaccines, at a value of $1.48 billion. These vaccines are used in routine immunization as well as in immunization campaigns.

Operations research (OR) is essential for developing a strong knowledge base and identifying innovative strategies to improve performance of health programs (Ryman and Dietz 2008). A few papers apply the mathematical programming tools to model healthcare problems in the field of vaccine supply management. Since the nature of VSCs is different from the commercial supply chains, it is not reasonable to directly apply the current models which have already been used in the commercial setting. The aim of this chapter is to identify and classify the gaps of vaccine supply problems which can help the scholars to fill such gaps using OR approaches and tools.

Table A.1 provides an overview of some VCS related issues which have potential to be modeled by mathematical programming tools (see Appendix A).

11.2 An Introduction to Vaccine Supply Chain Networks

Immunization programs play a key role in all over the world to prevent infectious and communicable diseases. To have a successful program, we need an effective while efficient vaccine supply chain.

A vaccine supply chain as an emergency supply chain has much in common with a commercial supply chain and it can be defined as "all personnel, systems, facilities, equipment, and activities involved in ensuring that vaccines are effectively delivered from the point of production to the people who need the vaccine" (Chopra and Meindl 2003).

Vaccine supply chain networks have several specific characteristics such as perishability, nonlinearity of health benefits of vaccination, non-profitability of vaccination programs for government as buyer and the value of averting infection dynamics (Chick et al. 2008) which differentiate them from commercial supply chains.

According to the rising cost of vaccines and the greater storage capacity required at every level of the related cold chain, World Health Organization (WHO) focuses on lower stock levels, reduce wastage, accurately forecast vaccine requirements, and prevent equipment break-downs. This requires a consistently high standard of supply chain management (WHO 2015b).

The role of VSCs is to ensure effective vaccine storage, handling, and stock management; rigorous temperature control in the cold chain; and maintenance of adequate logistics management information systems (WHO 2015c).

11.2.1 Vaccine Supply Chain Effectiveness and Efficiency

One of the main objectives of National Immunization Programs is to strengthen and optimize the immunization of VSCs so that vaccines are delivered to the end recipients effectively, efficiently and sustainably. An efficient VSC is essential to making the necessary vaccines available at the immunization locations for the population (Assi et al. 2013).

To have an effective and efficient supply chain of vaccine, it should meet several criteria including the product (i.e. providing correct product as per request), quantities (i.e. providing correct quantities as per request), place (i.e. immediately sending required vaccines to requested organization), time (i.e. timely supplying required vaccines), quality (i.e. providing required vaccines at appropriate conditions like right temperature) and costs (i.e. providing required vaccines at appropriate service charge) (Riewpaiboon et al. 2015; Zaffran et al. 2013).

Yadav et al. (2014) investigate the benefits and potential risks of integrating VSCs with other health commodities supply chains to improve the effectiveness and efficiency of the whole supply chain. They discuss horizontal supply chain integration in commercial sector and present some experiences about the integration of family products' planning with essential medicines. Different experiences in several countries such as Bolivia, Zambia, Nepal, Bangladesh and so on, show that supply chain integration of storage and distribution results in cost reduction. Considering quantification, procurement, storage in warehouses at the central, regional, and/or district levels (Foster et al. 2006; Milstien et al. 2005), they propose Fig. 11.1 as a VSC network and discuss the integration potential of each segment which results in savings in fixed and operating costs in areas related to transport, warehousing, distribution and supervisory control.Assi et al. (2013) investigate the impact of removing regional level from Niger VSC (Fig. 11.2) in order to improve its efficiency. In this paper, different cost indices are considered by which supply chain performance is measured. To evaluate different shipping policies, they focus on three key measures:

$$\textit{Vaccine availability} = \frac{\textit{Total number of vaccines available for children arriving over a one-year period}}{\textit{Total number of children arriving at a clinic over a one-year period}}$$

$$\textit{Transport capacity utilization} = \frac{\textit{Net liters of transport space consumed per shipment}}{\textit{Total liters of available transport space per shipment}}$$

$$\textit{Storage capacity utilization} = \frac{\textit{Net Liters of storage space consumed per time period}}{\textit{Total liters of available storage space per shipping interval}}$$

The shipping policies that are considered in this paper are as follows:

- Regional storage locations receive vaccines directly from the central storage locations when vaccines are available using a 4×4 truck as required.

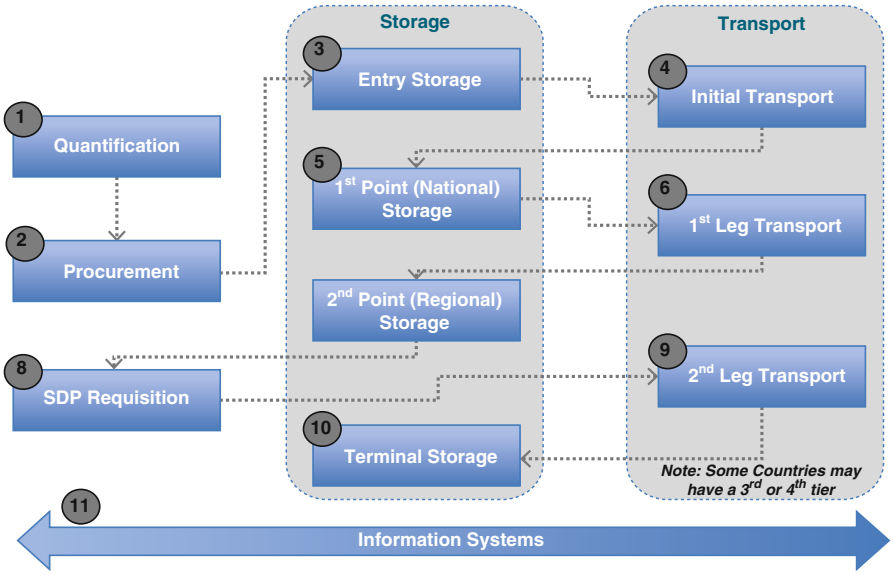

Fig. 11.1 Typical structure of a VSC (Yadav et al. 2014)

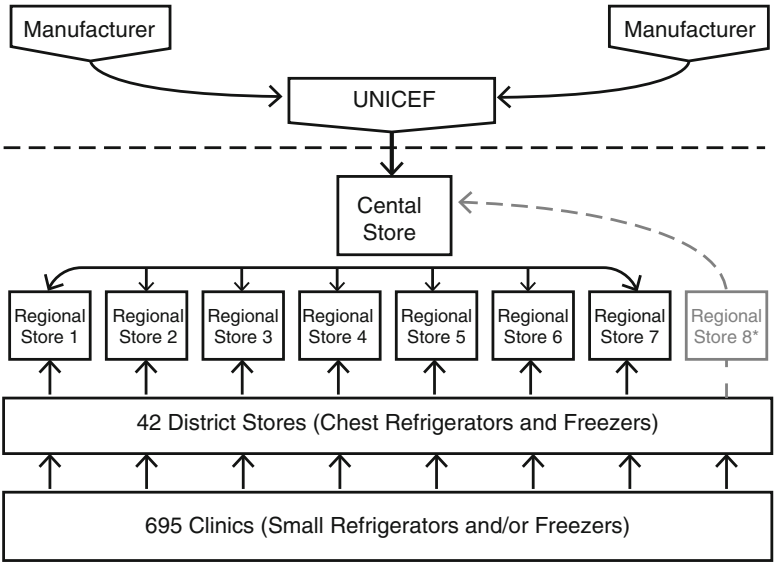

Fig. 11.2 Schematic of existing and alternative VSC frameworks in Niger (Assi et al. 2013). *Regional store 8 is out of service. Districts located in this region collect their vaccines directly from the central store

- The central storage location distributes vaccines directly to the regional storage locations in three cold trucks along three shipping loops.
- The central storage location distributes vaccines directly to the regional storage locations in eight cold trucks along eight shipping loops.

The results demonstrate that the performance and cost of the Niger VSC may improve by eliminating the regional level and making proper changes in shipping policies (i.e. collection-based shipments from the regions to the central level or monthly shipments from the central to regional storage locations).

Riewpaiboon et al. (2015) study the economic benefits of Vendor Managed Inventory (VMI) and outsourcing of VSC systems compared to conventional systems in optimizing the efficiency of Thailand immunization programs. They find that the VMI systems lead to decrease in the number of required logistics operators and logistical costs and improvement in the efficient use of resources using the information technology and a logistics agency.

11.2.2 Vaccine Supply Chain Costs

Cost-effectiveness analysis (CEA) is a form of economic analysis to compare the cost and outcomes of a course of action (Bleichrodt and Quiggin 1999). CEA is often used in the field of health services since it may be inappropriate to monetize the health effects. Decision-makers often use some methods such as CEA to evaluate available choices about health interventions (Weinstein et al. 1996). Therefore, estimating the costs of VSC and delivery services has a key role in decision making to evaluate the impact of new technologies and programmatic innovations in health services such as immunization systems.

There are several studies about VSC cost in different geographic areas such as Vietnam (US$0.24 per dose) (Mvundura et al. 2014), Mozambique (US$1.06 and US$1.21 per dose for two provinces) (VillageReach 2009), Benin (US$0.23 per dose) (Brown et al. 2014), Kenya and Tanzania (US$0.43 and US$0.88 per dose respectively) (Mvundura et al. 2015).

Mvundura et al. (2015) find out that the cost per dose of providing immunization services is higher than the supply chain costs. That is because of staff salaries which are the main portion of the service delivery costs. They investigate that in the supply chain costs, total costs per facility decrease as moving down in the supply chain, but the cost per dose delivered increase because of the fixed costs (Riewpaiboon et al. 2015). Hutton and Tediosi (2006) present a costing study to measure the incremental cost of adding a hypothetical malaria vaccine to the expanded program on immunization (EPI) schedule to enable estimation of cost-effectiveness of such a vaccine. This study presents the predicted cost per dose delivered and cost per fully immunized child (FIC) in Tanzania, which are key inputs to the CEA. According to their study, the vaccine delivery cost is obtained by the following formula:

$$\begin{aligned} &\textit{Vaccine delivery cost per dose} \\ &= \begin{pmatrix} \textit{purchase cost} + \textit{distribution cost} + \textit{delivery cost} \\ + \textit{training cost} + \textit{social mobilization cost} \end{pmatrix} \textit{per dose}. \end{aligned}$$

The authors also elaborate on analyzing and calculating each term of this equation. They suppose that the malaria vaccine requires three doses to fully immunize a child. Therefore, the cost per FIC is computed via multiplying the vaccine delivery cost per dose by three.

11.3 Vaccine Supply Problems

Although immunization is one of the great public health interventions, continued success depends on an available supply of the vaccines. Several experiences of vaccine supply disruptions shows that a fragile vaccine supply will be part of the immunization environment (Rodewald et al. 2006). Therefore, to handle such disruptions and have a successful immunization program, we should recognize the factors leading to the VSC problems.

In this section, the main challenges of vaccine procurement (i.e. supply) management are discussed through elaborating on the relevant issues including the "vaccine sourcing", "vaccine demand forecasting", "vaccine shortage", and "vaccine cold chain" in the following subsections.

11.3.1 Vaccine Sourcing

VSCs within humanitarian aid networks concern with the sourcing from different supply markets. Many researchers have introduced different quantitative and qualitative practical approaches for analyzing and selecting the sources of supplies from different markets, mainly based on price, quality, service, delivery, and management compatibility criteria (Stentoft Arlbjørn and Pazirandeh 2011).

The supplier selection problem in the context of vaccine procurement involves the assessment of commercial and technical suitability of prospective vaccine suppliers. The aim is to develop a list of reliable, reputable and technically competent suppliers of each vaccine. The selection of suppliers should be made in advance of any contracting process (WHO 2008). Therefore, selection of appropriate suppliers is a major requirement for an effective supply chain.

Ordoobadi (2009) presents several criteria including the delivery, responsiveness, assurance, quality and cost to help decision makers when dealing with selection of the best suppliers which is a major requirement to have an effective supply chain. For example, the Chiron plant in Liverpool had to be closed down because of violations of Food and Drug Administration (FDA) standards which resulted in the lack of

the United States supply of flu vaccines in the fall of 2004 (Federgruen and Yang 2008). There were numerous reports which have documented supply problems with influenza and other vaccines, therefore, the sudden elimination of one of only two manufacturers was an unforeseeable and rare event which lead to elimination of half of the national vaccine supply (Heinrich 2001a, b, c). Similar problems in vaccine sourcing lead to vulnerability of immunization programs.

Federgruen and Yang (2008) study the multi-sourcing problem with unreliable suppliers, to cover uncertain demand of a given item such as vaccine. They consider that each source faces a random yield factor with a general probability distribution on the unit interval. The objective of this problem is to selecting suppliers from the given set and determining how much to order from each supplier to minimize total procurement costs while ensuring that the stochastic demand is met with a given probability.

There are some countries that have to procure their required doses of vaccines globally. In these countries, there are two separate, often not well-connected, supply chains: (1) the segment that delivers vaccines from global suppliers to the receiving countries (i.e. international segment) and (2) the segment that moves vaccines from the port of entry through the national and local storage and distribution system to the health care provider (i.e. national segment). Two potential improvements in the international segment would be to identify better transportation methods for vaccines and to set up regional storage hubs to stock vaccines in preparation for a pandemic or an emergency (Kaufmann et al. 2011).

There are some criteria to source vaccine from the global market such as, on-time delivery, low price, unavailability of local market and better quality. Also, some problems in global sourcing may include currency fluctuation, quality assurance, transport delays and political instability (Stentoft Arlbjørn and Pazirandeh 2011).

11.3.2 Vaccine Demand Forecasting

Based on WHO (2015a) and UNICEF (2012), the purpose of demand forecasting is to ensure the availability of an adequate supply of high quality vaccines and to estimate the quantity of goods and financial needs necessary to conduct immunization programs. When a sudden fluctuation in vaccine demand is expected (for example, during the introduction of new vaccines into the EPI, at the beginning and end of accelerated immunization activities and during outbreaks) forecasting the demand and coordinating the supply capacity at the global level in advance is critical. Therefore, the accuracy of the forecast is important, because underestimating the requirements results in vaccine shortages and overestimating results in excess stock which increase the manufacturers' and accessibility costs. The value of the forecast depends on its accuracy (taking into consideration the type of vaccine, the presentation (vial size), the quantity and the timing of delivery of the vaccine). Poor forecasting may result in delays or shortfalls in delivery and additional costs. Noteworthy, Single-dose vaccine formats can prevent the vaccine wastage in the

clinical level but may incur higher production cost, medical waste disposal, and storage costs in comparison to the multi-dose formats.

Furthermore, the availability of quality vaccine could affect the decision-making process for disease control policies, especially when the global vaccine need is expected to exceed the global production capacity. It is essential that the demand growth rate is to be coordinated with scale-up activities of vaccine producers to avoid shortages. In particular, in times of disease outbreak response, the distribution of available resources must be well coordinated at the global level.

National vaccine forecasting is done using population estimates, birth rates, infant mortality rates, vaccine waste rates, and prior-year estimates of usage (Kaufmann et al. 2011).

Chiu and Kuo (2006) use Autoregressive Integrated Moving Average (ARIMA) and neural networks model respectively to anticipate the yearly total number of vaccinations and the annual birthrate for the next year to forecast the vaccine demand of the next year.

Amarasinghe et al. (2010) estimate the potential vaccine demand for the public and private sectors in disease endemic countries and of the travelers' market in non-endemic countries. There are some limitations in demand forecasting such as vaccine wastage estimation because of the vial presentation (single or multi-dose), estimation of vaccine coverage rates based on an imperfect model, little experience with administration of a multi-dose vaccine to older children in catch-up immunization programs, not accounting for the number of doses that may be needed to control dengue outbreaks and the issue of vaccine safety. Despite these limitations, they present initial estimates of dengue vaccine demand and different introduction scenarios which provide the first step to forecast the production capacities for manufacturers and vaccine availability for national policymakers and potential donor agencies. The following formula is used to calculate the number of dengue vaccine doses for the 5-year introduction period in each dengue-endemic country:

$$\begin{aligned} &\textit{number of doses} \\ &= (\textit{target population})^{*}\,(\textit{expected coverage}) \\ &^{*}\,(\textit{scheduled number of doses}) \,/\, (1 - \textit{wastage}) \end{aligned}$$

11.3.3 Vaccine Shortage

Sometimes, the amount of a certain vaccine cannot keep up with the number of people who need it. Vaccine shortages can result from higher-than-expected demand, interruptions in production/supply, or lack of enough resources to purchase required vaccines (Hinman et al. 2006). Dal Moro (2013) believes that only about 10% of shortages can be attributed to lack of raw material and essential ingredients for drug manufacturing. Instead, most shortages are because of the manufacturers'

Table 11.1 SWOT matrix of 2004–2005 vaccine shortage

	Strengths	Weaknesses	Opportunities	Threats
Production	New entry	Reliance on egg supply	Greater acceptance of nasal spray vaccine	Unstable demand
		Drawbacks to new technologies	Increased demand	Controversy over thimerosal
		Overemphasis on liability	Large government investment	
Purchasing and distribution	Newly established relationships	Limited role of MOUs	Purchasing from multiple manufacturers	Purchasing from abroad
				Private physician stocks
Provision	Incident command system	Primacy of individuals	Priority-group identification	Darwinian implementation
				Continued racial disparity
				Inappropriate risk models
				End-of-season surplus

Uscher-Pines et al. (2008)
MOUs memoranda of understanding

decisions to stop the production, or interruptions in production caused by money problems.

There are several steps in reducing the likelihood of vaccine shortages like increased governmental support for research and development, enhanced private-public partnerships and advance purchase agreements (Hinman et al. 2006).

Some impacts of vaccine shortages are prioritization of high-risk patients, changes in vaccination rates and clinical practice, vaccine redistribution and patient referral and future plans for vaccine delivery (Kempe et al. 2007).

Uscher-Pines et al. (2008) applied the Strength, Weakness, Opportunity and Threat (SWOT) analysis to vaccine shortage analysis, which was led to three policies shown in Table 11.1.

11.3.3.1 Vaccine Production

- Strengths

New entry: The entry of a new manufacturer such as GlaxoSmithKline (GSK) can help the US vaccine market to cope with the recent shortages.

- Weaknesses

Reliance on egg supply: Availability of the vaccine eggs play a key role in vaccine production and supply. Therefore, rapid influenza vaccine supply in the case of shortage is impossible since millions of eggs must be ordered in advance of the 6 to 8-month production process.

Drawbacks to new technologies: The manufacturers prefer to change their technologies into the cell- rather than egg-based production, which makes them not to increase the plant capacity based on the existing technology. However, this transitional stage actually diminishes the current capacity for response.

Overemphasis on liability: When there is an overemphasis on liability to lawsuits, the vaccine market becomes unattractive.

- Opportunities

Greater acceptance of nasal spray vaccine: Despite resisting to Med Immune's FluMist product in comparison to injectable vaccines, the positive effect of the nasal spray vaccine during the 2004–2005 shortage beyond the 5–49 years age group may enhance demand. As the public becomes more comfortable with alternative modes of vaccine delivery beyond the traditional injection, research into such technologies will increase.

Increased demand: Media attention on the 2004–2005 shortage and the evolving threat of pandemic influenza may increase awareness of demand for seasonal influenza vaccine.

Large government investment: Another opportunity of the production phase is the large government investment in influenza research, which has grown from $21 million in 2001 to an estimated $119 million in 2005.

- Threats

Unstable demand: The demand of vaccine is uncertain because it is usually associated with the severity of the influenza season. Therefore, the exact demand becomes apparent after the supply of the season is set.

Disagreement over thimerosal: Thimerosal remains very controversial due to the lack of evidence linking the mercury-containing vaccine preservative thimerosal and the increased prevalence of autism. Therefore, as a precautionary measure the public health service agencies, American Academy of Pediatrics and vaccine manufacturers decided to reduce or even eliminate the use of thimerosal in response to growing public concern. However, only limited quantities of thimerosal-free influenza vaccine are produced each year, and the influenza vaccine is the only recommended paediatric vaccine in the US market to contain the preservative. Continued objections against thimerosal could decrease demand for influenza vaccine, especially among parents of young children.

11.3.3.2 Purchasing and Distribution

- Strengths
- Newly established relationships: When the shortage of influenza vaccine occurred in 2004–2005, many local public health agencies developed partnership with other agencies such as nursing homes and non-traditional partners. Weaknesses

Limited role of memoranda of understanding (MOUs): Although planners continue to rely upon MOUs in influenza preparedness (by which institutions agree to share resources), such agreements are much more suitable to a localized shortage not for the case in which the entire country faces the same dire predicament.

- Opportunities

Purchasing from multiple manufacturers: One of the most important experiences from 2004 to 2005 shortage of influenza vaccine, was providing the required vaccines from several suppliers (i.e. relying on multi-sourcing policy) rather than one supplier.

- Threats

Purchasing from abroad: A threat in purchasing and distribution phase is to fill the gap of shortage by providing the unlicensed vaccine from abroad, which is in contrast with FDA regulation.

Private physician stocks: According to the reports from the 2005 to 2006 influenza season, it is observed that manufacturers would prefer the large vaccine orders of mega-stores such Costco and CVS,[1] in order to address the spot shortages at physicians' offices.

11.3.3.3 Provision

- Strengths

Incident command system: The 2004–2005 shortage of influenza vaccine urged the local health departments to organize influenza clinics and to manage distribution points using the incident command approach.

- Weaknesses

Primacy of individuals: public health emergencies may face with crisis if the clinical medicine turn to the needs of individuals.

[1]CVS Pharmacy is a subsidiary of the American retail and health care company CVS Health, headquartered in Woonsocket, Rhode Island. It was originally named the Consumer Value Store, and was founded in Lowell, Massachusetts in 1963.

- Opportunities

Priority-group identification: Prioritization of different groups plays a vital role in coping with the pandemic in the case of vaccine shortage. Therefore, an important lesson learned from the 2004 to 2005 shortage is to ration vaccine among prioritized groups.

- Threats

Darwinian implementation: In theory, the most vulnerable groups are in priority to receive immunization program such as vaccine which may include the young, the old and the chronically ill individuals. However, in the actual provision of vaccine in 2004–2005, a Darwinian "survival of the fittest" regime took hold which may further undermine public confidence in the health system's equity and fairness.

Continued racial disparity: One of the threats in the provision phase is the raising attention to the racial disparities in uptake of seasonal influenza vaccine.

Inappropriate risk models: New researches demonstrate that school-aged children are usually in the danger of epidemic and their immunization is the best way to protect adults. Therefore, using an inappropriate risk and priority model may lead to the waste of limited resources.

End-of-season surplus: There is usually a surplus of vaccine at the end of the influenza season even when there is a severe shortage and rationing at the beginning. In 2004–2005, public dismay regarding the lack of vaccine and initial restrictions on priority groups led to non-sustained demand; which finally resulted in three million doses unused in April 2005.

11.3.4 Vaccine Cold Chains

Cold chain is a system of different elements, i.e. human, material and financial resources, and certain norms and standards that ensure high-quality vaccines. This network includes refrigerators, cold stores, freezers and cold boxes to maintain vaccines at the right temperature during transportation, storage and distribution from manufacturer to the point of use.

There are several challenges in managing the vaccine cold chains including organizing the required cold chain, estimating the storage capacity, managing the equipment, planning and monitoring the cold chain (WHO 2004).

Accidental freezing is a serious threat that may occur in each phase of shipping the vaccines from the manufacturer to the point of use and leads to failure in immunization programs (Lloyd et al. 2015).

Matthias et al. (2007) classify the relevant papers in two different classes: (1) collecting data on temperature conditions in the existing cold chain infrastructures including transport and storage; (2) using temperature-measuring devices that register freezing temperatures. They compare several parameters such as sample size, freeze-exposure threshold, the year that the study was conducted, and the

Table 11.2 Capital expenditures required by level to achieve each scenario without bottlenecks

	National level ($)	Department level ($)	Commune level ($)	Health post level ($)	Total ($)
(a)	85,188	124,899	51,000	24,000	285,087
(b)	85,188	124,899	–	24,000	234,087
(c)	85,188	124,899	–	24,000	234,087
(d)	85,188	249,799	–	24,000	358,987

level of monitoring rigor under four different scenarios including developed versus developing country and transport versus storage of freeze-sensitive vaccines. Their analysis shows that the average proportion of exposure ranges from 14% to 35%. Finally, they present some effective tools to minimize the potential of vaccine freezing.

Brown et al. (2014) investigate the vaccine cold chain of Benin and develop a detailed, discrete-event simulation model to test the effects of different changes to current system on cost and vaccine availability. They consider several scenarios as shown in Fig. 11.3 and obtained the results summarized in Table 11.2.

11.4 Vaccine Supply Chain Coordination

Optimal supply chain performance requires the execution of a precise set of actions. These actions should be taken by cooperating of the supply chain members. But unfortunately supply chain members primarily are concerned about their own profit. Therefore, there should be some incentives that persuade them to coordinate in order to achieve an optimal performance for the whole chain (Cachon 2003).

Several contract models are discussed in the literature such as, the wholesale price only contract, buyback contract, revenue sharing contract, quantity flexibility contract and quantity discount contract. However, the contracts that coordinate the firms in a commercial supply chain cannot fully coordinate the VSCs members.

Many scholars investigate the market of influenza vaccine, since influenza is a respiratory illness that spreads rapidly and results in seasonal epidemics. Influenza occurs globally with an annual attack rate estimated at 5–10% in adults and 20–30% in children. Illnesses can result in hospitalization and death mainly among high-risk groups (including the very young, elderly or chronically ill). Worldwide, these annual epidemics are approximated to result in about 3–5 million cases of severe illness, and about 250,000–500,000 deaths (WHO 2014).

Vaccination is the most effective and efficient way to cope with the spread of disease and severe effects of infectious. On-time delivery of required vaccines has a key role in immunization program.

Chick et al. (2008) investigate different contracts and some features that enable cost sharing contract to coordinate a VSC considering yield uncertainty. They prove that the wholesale price contract in which the supplier (a single vaccine

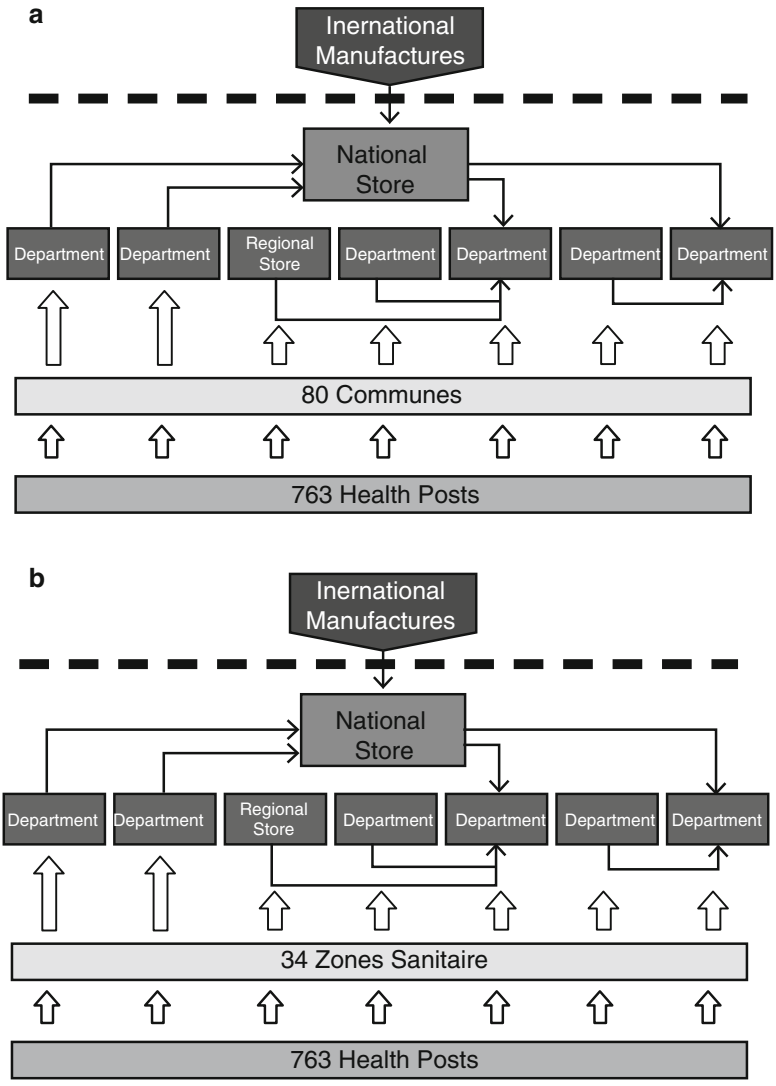

Fig. 11.3 Current state of VSC structure and restructured VSC. (**a**) Current Benin VSC structure. (**b**) Health Zone: For the 4-level Health Zone structure, the existing 80 Communes in the current system are fixed into 34 Health Zones. The Health Zone picks up the required vaccine from the Department/Region monthly and Health Posts pick-up from the Health Zones as needed. (**c**) Removing the Commune Level: This scenario considers a 3-level supply chain by removing the entire commune level instead of 4-level supply chain and the Health Posts directly pick up vaccine from the six Department Stores and the Regional Store. (**d**) Removing the Commune Level and Expanding to 12 Department Stores: This scenario also converts the supply chain into a 3-level structure by removing the commune level but also includes five additional Departmental Stores and converts the existing Regional Store to a Department Store for a total of 12 Department Stores. The National Depot delivers vaccines bimonthly to Departments in a loop, with each trip from the National Depot delivering to three Departments (Brown et al. 2014)

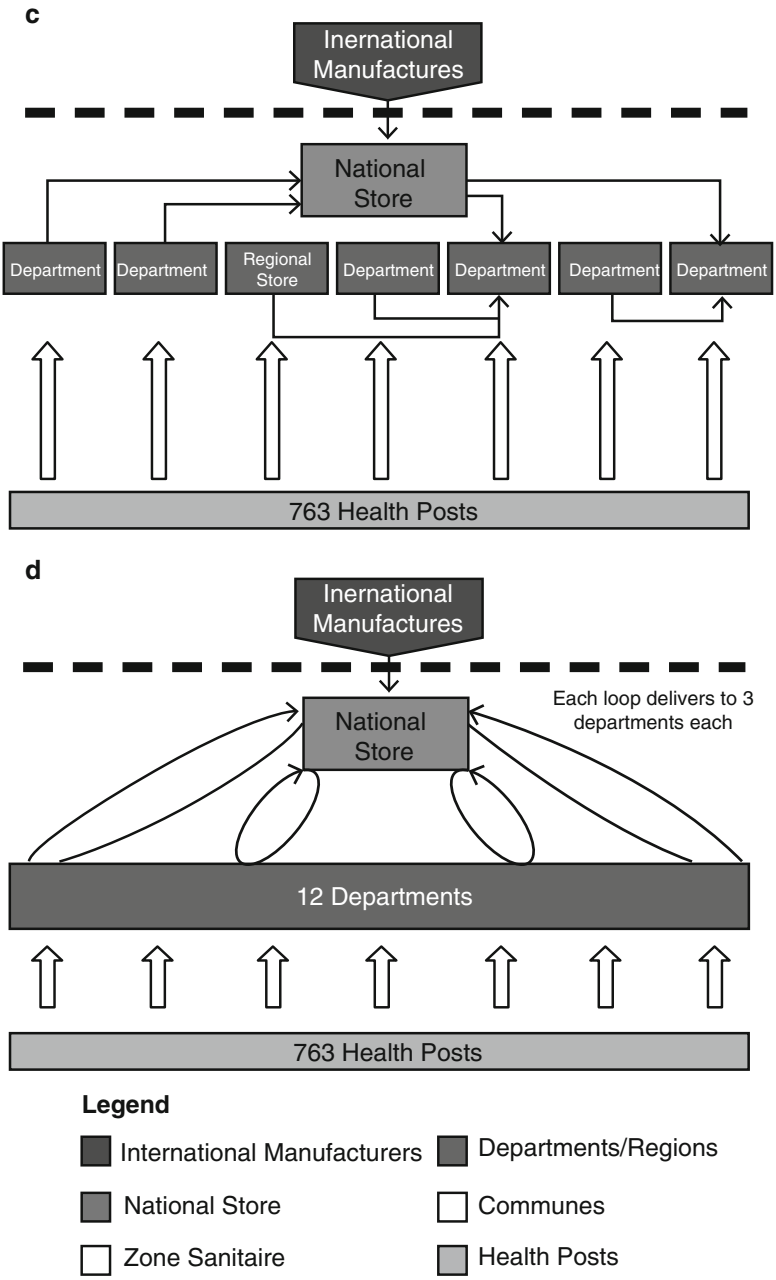

Fig. 11.3 (continued)

manufacturer) and the buyer (a single governmental public health sector) negotiate on a price, cannot fully coordinate the VSC. As well, they show that the pay back contract in which the government as the buyer agrees to buy any surplus production for a discounted price, does not provide enough incentive to coordinate the influenza supply chain while it does in commercial supply chains. They demonstrate a variant of the cost sharing contract that coordinates the influenza VSC in order to attain a global social optimum. Such an agreement provides an incentive to increase production since the risk of excess production by the manufacturer reduces. In the cost-sharing contract, the government pays an agreed part of the production level. As an example for application of OR tools in this context, we provide some details about the model developed by Chick et al. (2008). In this research, the manufacturer's problem is as follows:

$$\begin{array}{c} \min_{n_E} MF = E\left[cn_E - p_r Z\right] \text{ (Net manufacturer costs)} \\ s.t. Z = \min\left\{n_E U, fNd\right\} \text{ (Doses sold} \leqslant \text{Yield and demand)} \\ n_E \geq 0 \text{ (Non} - \text{negative production volume)} \end{array} \tag{11.1}$$

where U is random egg yield with $pdf\, g_U(u)$, and fNd is the government's vaccine order quantity. Therefore, the optimal production level n_E^* for the manufacturer is determined by

$$\int_0^{\frac{fNd}{n_E^*}} u g_U(u) du = \frac{c}{p_r} \tag{11.2}$$

where c is the unit cost of production per egg unit and p_r is the manufacturer's revenue per dose of vaccine from the government side. It is worth noting that the most common way for production of flu vaccines is using an egg-based manufacturing process. This production method requires large numbers of chicken eggs to produce vaccine and usually takes the longest period of time to produce vaccine. The government has to decide about the fraction f, while knowing that the manufacturer will behave optimally and may deliver less, in expectation, than what is ordered due to yield losses. The problem is then formulated as follows:

$$\begin{array}{c} \min_f GF = E\left[bT\left(\frac{W}{Nd}\right) + p_a W + p_r Z\right] \text{ (Net government costs)} \\ s.t. \quad Z = \min\left\{n_E U, fNd\right\} \text{ (Doses} \leqslant \text{yield and demand)} \end{array} \tag{11.3}$$

$W = \min\left\{n_E U, fNd, \bar{f}Nd\right\}$ (Doses given $\leqslant$doses purchased, cost effective level)
$\int_0^{\frac{fNd}{n_E^*}} u g_U(u) du = \frac{c}{p_r}$ (Manufacturer acts optimally)
$0 \leq f \leq 1$ (Fraction of population)
$n_E \geq 0$ (Nonnegative production level)

where b is the average total social cost per infected individual, $T(f)$ is the total expected number of infected people during the infection season, a function of the

fraction vaccinated, W is the number of doses administrated by government to susceptible population, N is the total number of population, d is the dose of vaccine needed per person, p_a is the unit cost per dose for government to administer vaccine and Z is the number of doses sold by the manufacturer to government. This is a two-player game which has a Nash equilibrium. Via procurement contracts, the sum of their expected financial and health costs may reduce to a level that is below the sum of costs if each player acts individually. So, the following problem has to be solved.

$$\min_{f,n_E} SF = E\left[bT\left(\frac{W}{Nd}\right) + p_a W + cn_E\right] \text{ (Total system costs)} \tag{11.4}$$

$s.t. W = \min\left\{n_E U,\ fNd,\ \bar{f}Nd\right\}$ (Doses given $\leqslant$doses bought, cost effective level)

$0 \leq f \leq 1$ (Fraction of population)

$n_E \geq 0$ (Nonnegative production level) ■

The lack of coordination leaves the manufacturer with production yield risks and results in the most needed countries and vaccine excess in the regions where they are not as effective, if the players act rationally (Lemmens et al. 2014; Mamani et al. 2010).

Decisions about the immunization against influenza in one country can affect the size of an outbreak in other countries due to interdependent risks from infectious disease transmission. Therefore, Mamani et al. (2013) propose a cost-sharing contract to reduce the total financial burden of an infection globally when considering multiple governments and the possibility of disease transmission across national boundaries, rather than focusing on a single buyer. The key idea is that the source country is subsidized to increase its vaccination effort by other countries that also benefit from that vaccination effort.

Arifoglu et al. (2012) study the effect of yield uncertainty in the supply side and self-interested consumers in demand side on the inefficiency of influenza VSC. Therefore, to achieve the coordination among the whole supply chain's members, there would be a need to combine demand-side interventions (such as tax-subsidy mechanisms as used by Brito et al. (1991)) and supply-side interventions (such as cost-sharing contracts as applied by Chick et al. (2008)).

Arifoglu et al. (2012) use backward induction to characterize the sub-game perfect equilibrium of the resulting two-stage game. In the second stage of the game, given the priority of groups and obtained number of doses, each individual makes his own vaccination decisions. By defining $\delta^{E}(\beta, Q_r)$ as the infection disutility of the marginal individual in equilibrium, the vaccine demand is given by $N\left(\overline{G}\left(\delta^{E}\right)\right)$ and the vaccinated fraction is calculated by:

$$h\left(\delta^{E}, UQ\right) = \min\left\{\overline{G}\left(\delta^{E}\right), \frac{Q_r}{N}\right\} \tag{11.5}$$

In the first stage, they present the manufacturer's problem by which the profit maximizing production quantity, Q, is determined as follow:

$$\max_{Q\geq 0}\pi_E(Q) = \text{ש}N\mathbb{E}_u\left[h\left(\delta^E, UQ\right)\right] - cQ \tag{11.6}$$

where U is the random yield, ש denotes the price of a dose of vaccine from manufacturer's side and c denotes the cost of each planned egg to the manufacturer. They show that the best value of the production quantity, Q_E, is determined by:

$$\text{ש}\int_0^{N\left[\overline{G}\left(\delta^E\right)\right]/Q_E} udM(u) = c. \tag{11.7}$$

They investigate the centralized system and its optimal solution to find out the sources of inefficiencies when there is no coordination. Therefore, they consider the coordination of production and vaccination by a social planner (i.e. the government) in the centralized system to maximize the total social welfare. Thus, in the second stage of the game, the social planner's problem is:

$$\max_{\delta\in\left[0,\overline{\delta}\right]}\{NW\left(\delta,\beta,Q_r\right) + R\left(\delta,Q_r\right)\} \tag{11.8}$$

The solution of this problem determines the optimal egg production, Q, which maximizes the total expected social welfare by considering the priority groups β. Also, according to the calculated number of doses, Q_r, the planner determines the demand level, which maximizes the ex post total social welfare whose utility per individual is given by:

$$\begin{aligned} W\left(\delta,\beta,Q_r\right) = \overline{V} - \theta\left[\overline{G}\left(\delta\right)\right] - rh\left(\delta,Q_r\right) - p\left(h\left(\delta,Q_r\right)\right)\int_0^{\delta} z dG(z) \\ - p\left(h\left(\delta,Q_r\right)\right)\times\left[\left(1-\phi\left(\max\left\{\delta,\beta\right\},\delta\right)\right)\int_{\max\{\beta,\delta\}}^{\overline{\delta}} z dG(z)\right. \\ \left. + \left(1-\phi\left(\min\left\{\delta,\beta\right\},\delta\right)\int_{\min\{\beta,\delta\}}^{\overline{\delta}} z dG(z)\right]\right. \end{aligned} \tag{11.9}$$

And the total sales revenue is calculated as:

$$R\left(\delta,Q_r\right) = \text{ש}Nh\left(\delta,Q_r\right) \tag{11.10}$$

Also, in the first stage of the game, the problem of the social planner is given by:

$$\max_{Q\geq 0}\mathcal{W}_C(Q) = N\mathbb{E}_u\left[W\left(\delta^C(UQ),\beta,UQ\right)\right] + \mathbb{E}_u\left[R\left(\delta^C(UQ),\beta,UQ\right)\right] - cQ \tag{11.11}$$

They characterize the best value of the production amount in the centralized system by the following proposition:

(i) The best value of the production amount for the centralized system (Q_C) is unique as follows:

$$\int_0^{N\left[\overline{G}\left(\delta_1^C\right)\right]/Q_C} B\left(\delta_2^C\left(uQ_C\right)\right) dM(u) = c \tag{11.12}$$

where $\delta_2^C\left(Q_r\right) = \overline{G}^{-1}\left(\min\left\{1, \frac{Q_r}{N}\right\}\right)$ and $B(\delta)$ is given by:

$$B\left(\delta\right) \triangleq -p'\left(\overline{G}\left(\delta\right)\right) \int_0^{\delta} z dG(z) + \delta p\left(\overline{G}\left(\delta\right)\right) - \left(r + \theta - \omega\right). \tag{11.13}$$

(ii) The best value of the production amount in the centralized system (Q_C) is greater than that of in the decentralized system (Q_E) ■.

Adida et al. (2013) investigate the effects of subsidy programs as a coordinating mechanism to align the incentives of different modules in a vaccine market and present a subsidy scheme to coordinate the market. They show that a fixed one- or two-parameter subsidy structure cannot provide sufficient incentive to the manufacturer and the consumers to guarantee socially optimal vaccine coverage, but a two-part menu of subsidies can. They consider a two-parameter incentive mechanism in which a subsidy is provided to the consumers to coordinate the vaccine price (vaccine coverage) and, under a cost sharing contract, the central planner agrees to pay to the manufacturer for every production unit to coordinate production quantity.

Dai et al. (2014) consider three key sources of uncertainties in the US influenza VSC: (1) The product design is exogenous to a manufacturer because the Vaccine and Related Biologic Products Advisory Committee determines the composition of influenza vaccine; (2) The delivery lead time required for manufacturing and distributing vaccines is long (usually 6–8 months) and uncertain. (3) The demand is time-sensitive and uncertain. Despite some other studies that consider the effectiveness of potential government interventions through partial centralization or subsidies, they focus on improvement opportunities through coordinating contracts between firms in this supply chain. They analyze three contracts, namely, wholesale price, Delivery-time-dependent Quantity Flexibility (D-QF) and Late-Rebate (LR) contracts and proceed to show that D-QF contract, a contract adopted in practice, may not coordinate the supply chain due to the tension between overcoming double marginalization and incentivizing early production. Moreover, another contract used in practice, the LR contract, only coordinates the supply chain under unrealistic conditions, but nearly coordinates the supply chain when demand uncertainty is low. After that, they construct a Buyback-and-Late-Rebate contract to coordinate the VSC and provide flexibility of profit division.

11.5 Stockpiling Vaccine Supplies for Pandemics

As mentioned before, immunization program is an effective way to control the spread of infectious diseases especially in the case of a disaster. Therefore, inventory control to manage specific medical supplies like vaccines, may be so helpful to reduce the impact of an epidemic outbreak. In some cases, the problem of stockpiling of medical supplies has been treated as a common inventory stockpiling problem for several groups of hospitals, because one of the best ways to manage disruptions in vaccination is stockpiling vaccine supplies (Dasaklis et al. 2012) as it is also the case for prepositioning other relief supplies to response to a disaster immediately at early post-disaster phase (see for instance Tofighi et al. (2016)).

Vaccine pre-positioning (i.e. stockpiling) is a major part of the response plan to epidemic disease outbreaks for several decades, and opportunities for their use are likely to increase. Therefore, beyond preparation of a critical vaccine supply to respond to epidemic disease outbreaks, construction of stockpiles can drive innovations for the development of vaccines with longer shelf lives, increased efficacy, and more efficient delivery (Yen et al. 2015).

In the case of vaccine inventory control, Liu (2007) considers the case in which a sudden demand for vaccines occurs because of a disaster like a natural outbreak or a bioterrorist attack. Based on this study, it is necessary to design the network of vaccine stockpiles at most crowded regions to overcome potential shortage during a disaster. He develops two mathematical models (an over-simplified inventory model for emergency, and a more realistic model) for estimating the necessary inventory levels of vaccines in order to meet likely future urgent needs.

DeLaurentis et al. (2008) analyze the problem of determining the inventory level of one crucial medical item considering the uncertain demand that may occur under various possible pandemic scenarios. Also, DeLaurentis et al. (2009) investigate the problem of hospital stockpiling of crucial medical supplies in preparation for a possible influenza pandemic. They consider a regional network of hospitals that have mutual aid agreements in place such that they may borrow or lend supplies from each other during medical emergencies. Each hospital aims to minimize the expected total cost incurred by stockpiling medical supplies. It includes the purchasing cost of supplies, the holding cost until the start of the pandemic and the cost of borrowing required excess supplies from other hospitals in the network. Moreover, the hospitals get penalized when facing with supply shortage, and for setting low pre-determined limit. Therefore, a game theoretical approach is tailored to make the stockpiling decisions of a network of hospitals in anticipation of a flu pandemic.

Dhankhar et al. (2009) study the economic feasibility of stockpiling by economic evaluation to guide the policy makers. They investigate the required amount of vaccines to be held as an inventory. Also, they examine some economic analysis according to the vaccine's expiration date and vaccine control strategies in the case of secondary bacterial infections (Dhankhar et al. 2010).

Tebbens et al. (2010) determine the optimal control policies of a vaccine stockpile using a mathematical model for eradicated diseases like polio over time while accounting for several constraints such as capacity constraints, production and filling delays and the expiry process, risks associated with the stockpiling, uncertainty and dynamics of vaccine demands, problems of investment, location, and the implications of possible changes over the time. Via solving such a mathematical model, the optimal ordering strategy is determined aiming at the minimization of the present value of public health and vaccine costs.

Optimizing the vaccine stockpile involves making a trade-off between vaccine costs and public health costs. They model two different aspects of the optimization problem: Aspect 1: Minimize the present value of total costs over all feasible ordering strategies, assuming no financial constraints:

$$Min \;\; C = \int_{t_0}^{\infty} \left(c_p(t) + c_v(t)\right) e^{-rt} dt \tag{11.14}$$

$$s.t. o(t) \geq 0 \tag{11.15}$$

where C is the net present value of the total costs, c_p and c_v are the order rate dependent annual public health costs and annual vaccine costs, respectively, r is the discount rate, and o is the order rate. The constraint shows that any ordering strategy that leads to nonnegative order rate is feasible. If the order rate exceeds the maximum production rate, excess orders will simply remain in the stock vaccine in production as backlog and the production constraints will govern the rate at which these orders ultimately flow to the stockpile. Aspect 2: Minimize the present value of public health costs over all feasible ordering strategies subject to the financial constraints:

$$Min \;\; C_p = \int_{t_0}^{\infty} c_p(t) e^{-rt} dt \tag{11.16}$$

$$s.t. o(t) \geq 0 \tag{11.17}$$

$$0 \leq \int_{t_0}^{\infty} c_v(t) e^{-rt} dt \leq F \tag{11.18}$$

where C_p is the net present value of the public health costs, F represents the financial constraints as the total amount of available funds (alternatively, the financial constraint could consist of variable maximum funds available over the planning horizon). Under aspect 2, the available funds determine the lowest possible health costs. While, under aspect 1, the potential public health costs determine the amount of funds available for the stockpile.

Yen et al. (2015) explore the related literature of vaccine stockpiling and address some general problems and complexity of decision about establishing and maintenance of a vaccine stockpile. Also, they investigate the solutions of stockpile management, funding, and ethical issues regarding access to vaccines and how the process needs to account for disease and vaccine characteristics.

11.6 Conclusion Remarks

As noted before, there are many causes for occurrence of pandemic outbreaks such as natural or man-made disasters. Most of such diseases are preventable by vaccination plans. “Prevention is better than cure”, therefore, vaccination plays a key role in all over the world to prevent infectious and communicable diseases and it is a valuable way to do so. Sooner prevention (by vaccination) is more effective, but timely control of an epidemic outbreak strongly depends on the pre-establishment of an emergency supply chain network adopted at international, national or community level. Designing an appropriate VSC network is an important issue which differs from commercial supply chain networks because of some special characteristics. Also, improving the efficiency and effectiveness of VSC has a key role in a successful immunization program.

In this chapter, we investigated the vaccine supply chains regarding some related challenges. There are many research gaps in this area that need to be addressed from the operations management and optimization point of view. The main problems of VSCs include vaccine sourcing, vaccine demand forecasting, vaccine shortages and the cold chain that suffer from the lack of mathematical modeling and precise solutions. Solving such problems may help the responsible public health organizations to improve the immunization coverage. Further researches may include developing a mathematical model for forecasting the vaccine demand of the affected areas which has a vital role in controlling the epidemic diseases and lowering the social cost caused by such epidemics. In addition, vaccine sourcing from international oligopoly markets requires an appropriate coordination contract and pricing strategy.

Also, on-time delivery of required vaccines which has a key role in an immunization program can be reached by VSC coordination and appropriate stockpiling of necessary vaccines and need more attention from mathematical programming perspective to obtain accurate results. More importantly, stockpiling of vaccine supplies is an effective way in disease control interventions especially in the case of disasters and pandemics.

Appendix A

Table A.1 An overview of some VCS related issues

		Research	Year	Description	Application of OR tools in the paper
VSC effectiveness and efficiency		Assi et al.	(2013)	Effective and efficient VSC	✗
		Zaffran et al.	(2013)		✗
		Riewpaiboon et al.	(2015)	Benefits of VMI and VSC outsourcing	✗
		Yadav et al.	(2014)	Benefits/risks of integrating VSC	✗
		Mvundura et al.	(2015)	Cost of VSC and immunization service providing	✗
		Hutton and Tediosi	(2006)		✗
Vaccine supply problems	Vaccine sourcing	Stentoft Arlbjørn and Pazirandeh	(2011)	Humanitarian aid sourcing from different markets	✗
		Ordoobadi	(2009)	Supplier selection	✗
		Federgruen and Yang	(2008)	Multi-sourcing problem with unreliable suppliers	✓
		Kaufmann et al.	(2011)	Problems of global vaccine sourcing	✗
	Vaccine demand forecasting	Chiu and Kuo	(2006)	Total yearly demand forecasting using ARIMA and neural networks	✓

(continued)

Table A.1 (continued)

		Research	Year	Description	Application of OR tools in the paper
		Amarasinghe et al.	(2010)	Forecasting the potential vaccine demand in endemic countries and of the travelers market in non-endemic countries	✗
	Vaccine shortage	Hinman et al.	(2006)	Steps of reducing the likelihood of vaccine shortages	✗
		Kempe et al.	(2007)	Impacts of vaccine shortages	✗
		Uscher-Pines et al.	(2008)	Applying SWOT analysis to vaccine shortage	✗
	Vaccine cold chain	Lloyd et al.	(2015)	Failure in immunization due to accidental freezing	✗
		Matthias et al.	(2007)	Review of relevant papers	✗
		Brown et al.	(2014)	Investigating the vaccine cold chain of Benin	✗
Vaccine supply chain coordination		Chick et al.	(2008)	Influenza VSC coordination using a variant of the cost sharing contract	✓
		Mamani et al.	(2013)	Cost-sharing contract to decrease global financial burden of disease	✓
		Arifoglu et al.	(2012)	Combining demand-side and supply-side interventions to coordinate the entire supply chain	✓

	Adida et al.	(2013)	Investigate the effects of subsidy programs as a coordinating mechanism	✓
	Dai et al.	(2014)	Analyzing wholesale, D-QF and LR contracts in VSC coordination	✓
Stockpiling vaccine supplies for pandemic	Delaurentis et al.	(2008)	Determining stockpile quantity of one critical medical item	✓
	Delaurentis et al.	(2009)	Hospital stockpiling of critical medical supplies	✓
	Dhankhar et al.	(2009)	Studying economic feasibility of stockpiling	✗
	Tebbens et al.	(2010)	Determining the optimal management of a vaccine stockpile	✓

References

Adida E, Dey D, Mamani H (2013) Operational issues and network effects in vaccine markets. Eur J Oper Res 231(2):414–427

Amarasinghe A, Wichmann O, Margolis HS, Mahoney RT (2010) Forecasting dengue vaccine demand in disease endemic and non-endemic countries. Hum Vaccin 6(9):745–753

Arifoglu K, Deo S, Iravani SMR (2012) Consumption externality and yield uncertainty in the influenza vaccine supply chain: Interventions in demand and supply sides. Manag Sci 58(6):1072–1091

Assi T-M, Brown ST, Kone S, Norman BA, Djibo A, Connor DL, Wateska AR, Rajgopal J, Slayton RB, Lee BY (2013) Removing the regional level from the Niger vaccine supply chain. Vaccine 31(26):2828–2834

Bleichrodt H, Quiggin J (1999) Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost–benefit analysis? J Health Econ 18(6):681–708

Brito DL, Sheshinski E, Intriligator MD (1991) Externalities and compulsary vaccinations. J Public Econ 45(1):69–90

Brown ST, Schreiber B, Cakouros BE, Wateska AR, Dicko HM, Connor DL, Jaillard P, Mvundura M, Norman BA, Levin C (2014) The benefits of redesigning Benin's vaccine supply chain. Vaccine 32(32):4097–4103

Cachon GP (2003) Supply chain coordination with contracts. Handbooks Oper Res Manag Sci 11:227–339

Chick SE, Mamani H, Simchi-Levi D (2008) Supply chain coordination and influenza vaccination. Oper Res 56(6):1493–1506

Chiu R-K, Kuo C-W (2006) A decision support model of demand forecast for national immunisation vaccines. Int J Electron Healthc 2(1):92–104

Chopra S, Meindl P (2003) Supply chain management: strategy, planning and operation. Pearson Education, Inc., Upper Saddle River

Dai T, Cho S-H, Zhang F (2014) Contracting for on-time delivery in the US influenza vaccine supply chain. Available at SSRN 2178157

Dal Moro F (2013) BCG shortage in Europe. Prev Med 2(57):146

Dasaklis TK, Pappis CP, Rachaniotis NP (2012) Epidemics control and logistics operations: A review. Int J Prod Econ 139(2):393–410

DeLaurentis P-C, Adida E, Lawley M (2008) A game theoretical approach for hospital stockpile in preparation for pandemics. Hospital 101:1

DeLaurentis P-CC, Adida E, Lawley M (2009) Hospital stockpiling for influenza pandemics with pre-determined response levels. In: Service operations, logistics and informatics, 2009. SOLI'09. IEEE/INFORMS international conference on: IEEE

Dhankhar P, Dasbach EJ, Elbasha EH (2009) Economics of stockpiling for an influenza pandemic. Lancet Infect Dis 9(8):459–460

Dhankhar P, Grabenstein JD, O'Brien MA, Dasbach EJ (2010) Cost-effectiveness of stockpiling 23-valent pneumococcal polysaccharide vaccine to prevent secondary pneumococcal infections among a high-risk population in the united states during an influenza pandemic. Clin Ther 32(8):1501–1516

Federgruen A, Yang N (2008) Selecting a portfolio of suppliers under demand and supply risks. Oper Res 56(4):916–936

Foster S, Laing R, Melgaard B, Zaffran M (2006) Ensuring supplies of appropriate drugs and vaccines. World Bank, Washington, DC

Heinrich J (2001a) Flu vaccine: supply problems heighten need to ensure access for high-risk people: GAO-01-624, 15 May. US General Accounting Office, Washington, DC

Heinrich J (2001b) Flu vaccine: recent supply shortages underscore ongoing challenges. Testimony before the Subcommittee on Health and the Subcommittee on Oversight and Investigations, Committee on Energy and Commerce of the House of Representatives November 18

Heinrich J (2001c) Flu vaccine: steps are needed to better prepare for possible future shortages. US General Accounting Office, Washington, DC

Hinman AR, Orenstein WA, Santoli JM, Rodewald LE, Cochi SL (2006) Vaccine shortages: History, impact, and prospects for the future*. Annu Rev Public Health 27:235–259

Hutton G, Tediosi F (2006) The costs of introducing a malaria vaccine through the expanded program on immunization in Tanzania. Am J Trop Med Hyg 75(2 suppl):119–130

Kaufmann JR, Miller R, Cheyne J (2011) Vaccine supply chains need to be better funded and strengthened, or lives will be at risk. Health Aff 30(6):1113–1121

Kempe A, Daley MF, Stokley S, Crane LA, Beaty BL, Barrow J, Babbel C, Miriam Dickinson L, Steiner JF, Berman S (2007) Impact of a severe influenza vaccine shortage on primary care practice. Am J Prev Med 33(6):486–491

Lemmens S, Decouttere C, Vandaele N, Bernuzzi M (2014) Integrated supply chain network design models for vaccines: a literature review. FEB Research Report KBI_1414

Liu Y (2007) Mathematical models of vaccine inventory design for a breakout of epidemic disease. PAMM 7(1):2150013–2150014

Lloyd J, Lydon P, Ouhichi R, Zaffran M (2015) Reducing the loss of vaccines from accidental freezing in the cold chain: The experience of continuous temperature monitoring in Tunisia. Vaccine 33(7):902–907

Mamani H, Chick SE, Simchi-Levi D (2010) Influenza Vaccine Supply Chain with Multiple Agencies. Plant Biotechnol J 8:607–619

Mamani H, Chick SE, Simchi-Levi D (2013) A game-theoretic model of international influenza vaccination coordination. Manag Sci 59(7):1650–1670

Matthias DM, Robertson J, Garrison MM, Newland S, Nelson C (2007) Freezing temperatures in the vaccine cold chain: a systematic literature review. Vaccine 25(20):3980–3986

Milstien J, Batson A, Wertheimer AI (2005) Vaccines and drugs: characteristics of their use to meet public health goals. The World Bank Health, Nutrition and Population (HNP) discussion paper

Mvundura M, Kien VD, Nga NT, Robertson J, Van Cuong N, Tung HT, Hong DT, Levin C (2014) How much does it cost to get a dose of vaccine to the service delivery location? Empirical evidence from Vietnam's Expanded Program on Immunization. Vaccine 32(7):834–838

Mvundura M, Lorenson K, Chweya A, Kigadye R, Bartholomew K, Mohammed M, Patrick Lennon T, Mwangi S, Kirika L, Kamau P (2015) Estimating the costs of the vaccine supply chain and service delivery for selected districts in Kenya and Tanzania. Vaccine 33(23):2697–2703

Ordoobadi SM (2009) Development of a supplier selection model using fuzzy logic. Supply Chain Manag: Int J 14(4):314–327

Riewpaiboon A, Sooksriwong C, Chaiyakunapruk N, Tharmaphornpilas P, Techathawat S, Rookkapan K, Rasdjarmrearnsook A, Suraratdecha C (2015) Optimizing national immunization program supply chain management in Thailand: an economic analysis. Public Health 129:899–906

Rodewald LE, Orenstein WA, Mason DD, Cochi SL (2006) Vaccine supply problems: a perspective of the Centers for Disease Control and Prevention. Clin Infect Dis 42(Supplement 3):S104–S110

Ryman T, Dietz V (2008) The need to improve quality, rigour and dissemination of operations research. Bull World Health Organ 86(11):A–B

Stentoft Arlbjørn J, Pazirandeh A (2011) Sourcing in global health supply chains for developing countries: literature review and a decision making framework. Int J Phys Distrib Logist Manag 41(4):364–384

Tebbens RJD, Pallansch MA, Alexander JP, Thompson KM (2010) Optimal vaccine stockpile design for an eradicated disease: Application to polio. Vaccine 28(26):4312–4327

Tofighi S, Torabi SA, Mansouri SA (2016) Humanitarian logistics network design under mixed uncertainty. Eur J Oper Res 250(1):239–250

UNICEF (2012) Supplies and logistics (Vaccine forecasting). Retrieved 06.10.2015

Uscher-Pines L, Barnett DJ, Sapsin JW, Bishai DM, Balicer RD (2008) A systematic analysis of influenza vaccine shortage policies. Public Health 122(2):183–191

VillageReach (2009) Comparison of costs incurred in dedicated and diffused vaccine logistics systems: cost-effectiveness of vaccine logistics in Cabo Delgado and Niassa Provinces, Mozambique. VillageReach, Seattle

Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB (1996) Cost-effectiveness in health and medicine, vol 55. Oxford University, New York

WHO (2004) Mid-level management course for EPI managers, module: cold chain management: World Health Organization (WHO)

WHO (2008) Immunization supply and procurement (The cycle of vaccine procurement) Retrieved 06.10.2015

WHO (2014) Media centre (Influenza-seasonal). Retrieved 27.10.2015

WHO (2015b) Immunization, vaccines and biologicals (Effective Vaccine Management (EVM) Initiative). Retrieved 06.10.2015

WHO (2015c) Immunization, vaccines and biologicals (Immunization supply chain and logistics). Retrieved 06.10.2015

WHO (2015a) Immunization supply and procurement (Demand forecasting for routine immunization). Retrieved 06.10.2015

Yadav P, Lydon P, Oswald J, Dicko M, Zaffran M (2014) Integration of vaccine supply chains with other health commodity supply chains: a framework for decision making. Vaccine 32(50):6725–6732

Yen C, Hyde TB, Costa AJ, Fernandez K, Tam JS, Hugonnet S, Huvos AM, Duclos P, Dietz VJ, Burkholder BT (2015) The development of global vaccine stockpiles. Lancet Infect Dis 15(3):340–347

Zaffran M, Vandelaer J, Kristensen D, Melgaard B, Prashant Y, Antwi-Agyei KO, Lasher H (2013) The imperative for stronger vaccine supply and logistics systems. Vaccine 31:B73–B80

Part V
Care Process: Diagnosis and Prognosis

Chapter 12
OR Applications in Disease Screening

Evrim Didem Güneş and E. Lerzan Örmeci

12.1 Introduction

This chapter presents an overview of disease screening problems and operations research (OR) applications on different aspects of the problem. Prevention activities in healthcare can be classified into three categories. *Primary Prevention* aims at preventing diseases before they occur by making healthy people less vulnerable to diseases. Health promotion activities such as smoking cessation campaigns are in this category. *Secondary Prevention*, or *early detection* aims to detect a disease early, if possible at asymptomatic or preclinical stage. Screening for a disease is in this category. Finally, *tertiary prevention* aims to soften the impact of existing diseases, for example by preventing complications due to chronic diseases such as diabetes (Heidenberger 1996). In this chapter we focus on screening for cancer to illustrate a form of secondary prevention activity.

The topic has attracted significant attention from the OR community in recent years, and a number of recent book chapters and tutorials are available. See Steimle and Denton (2016) for a tutorial on Markov Decision Process (MDP) models and partially observed Markov Decision Process (POMDP) models for improving screening and treatment of chronic diseases, Alagoz (2014) for a tutorial on POMDP models for optimization of screening decisions, and Zhang et al. (2013) for a recent chapter on disease prevention, detection and treatment.

E.D. Güneş (✉)
College of Administrative Sciences and Economics, Koç University, 34450 İstanbul, Turkey
e-mail: egunes@ku.edu.tr

E.L. Örmeci
Department of Industrial Engineering, Koç University, 34450 İstanbul, Turkey
e-mail: lormeci@ku.edu.tr

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_12

Screening is a common practice in most developed countries for some cancers (breast, colon, lung, cervical, ovarian etc.) and for other chronic conditions such as Type-2 diabetes, hypertension, or for infectious diseases such as influenza, or HIV. In this chapter, we will be using colorectal cancer (CRC) screening problem as an example to illustrate the issues discussed throughout the chapter. Screening recommendations are made based on extensive research on the candidate diseases, published mostly in medical journals. In this review, we will concentrate on the research done by the OR community and mainly published in journals with OR focus. As a result, we aim to highlight possible research directions for the OR community. We discuss different modeling techniques used to analyze different aspects of the problem, including both medical decision making and service management related issues.

In the remainder of this section we discuss the strategic decisions for screening from a health policy maker's perspective. In Sect. 12.2, we discuss OR applications in evaluation and optimization of screening policies. In Sect. 12.3 management of screening service delivery for reaching out to the population is discussed. Main topics included are location and resource allocation problems. We conclude with a discussion of future research directions in Sect. 12.4.

12.1.1 Assessing Population Level Screening

Screening for a disease is a strategic investment, requiring a significant amount of resources and involving large population segments. The decision to implement a screening program should be taken after careful consideration of the costs and benefits of such programs. A report commissioned by the World Health Organization defined the criteria that should be considered in making this important decision (Wilson and Jungner 1968). Since then, these criteria have been considered the gold standard in making such decisions (Andermann et al. 2008). The original criteria are given in Table 12.1. Evaluation of screening for CRC against these criteria reveals that most of the criteria are met.

Heidenberger (1996) provides an overview of approaches for screening program evaluation, and proposes the cost effectiveness analysis as the preferred framework in practice. Cost Effectiveness analysis (CEA) measures the effectiveness of programmes in terms of unit cost per an additional unit of health benefit measure. The most common of these measures is Quality Adjusted Life Years (QALY) gained. The number of QALYs is obtained by multiplying the life expectancy with a numerical weight between 0 (indicating death) and 1 (indicating perfect health) for the health state. Further, the total cost of the screening program is divided by the total number of QALYs gained relative to doing nothing, to find the cost-effectiveness ratio. If another policy is used as a benchmark, the ratio is called Incremental Cost Effectiveness Ratio. A cost-effectiveness ratio threshold of $ 50,000–$ 100,000 per QALY is frequently used for screening decisions, although this threshold may change over time and locations (Grosse 2008). OR methods such as simulation

Table 12.1 Criteria to evaluate population level screening programs (Wilson and Jungner 1968)

1. The condition sought should be an important health problem.
2. There should be an accepted treatment for patients with recognized disease.
3. Facilities for diagnosis and treatment should be available.
4. There should be a recognizable latent or early symptomatic stage.
5. There should be a suitable test or examination.
6. The test should be acceptable to the population.
7. The natural history of the condition, including development from latent to declared disease, should be adequately understood.
8. There should be an agreed policy on whom to treat as patients.
9. The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.
10. Case-finding should be a continuing process and not a once and for all project.

and stochastic models are widely used to estimate the cost effectiveness ratios for screening programs.

A screening program is mainly defined by the screening test to be used, target population in terms of the risk and age groups, and timing of consecutive screening tests. These parameters can be static as well as dynamic, depending on the age or the previous test result. Cost and accuracy of the screening modalities are important factors for cost effectiveness decisions. Other important factors pertain to the natural trajectory of the disease (in particular the progression rate), which influences the window of opportunity for early detection as well as the risk of overdiagnosis and overtreatment. In addition, prevalence of the disease in the population influences the potential benefits of population level screening.

One aspect of evaluating screening policies is to evaluate and compare different screening methods, which may differ in accuracy and cost as well as convenience. Screening test accuracy is measured by its sensitivity, i.e., the ability to detect existing disease (resulting in low false-negatives) and specificity, i.e. the ability of not being suspicious of a disease when the disease does not exist (resulting in low false positives). For example, in CRC screening, a number of different screening tests are available, such as fecal occult blood test (FOBT), double-contrast barium enema, sigmoidoscopy, computed tomography (CT) or virtual colonography and colonoscopy, with varying values of sensitivity and specificity as well as different levels of invasiveness and costs. Within these, colonoscopy is referred to as the gold standard test for CRC with the highest sensitivity (95% for cancer) and a very high specificity (90%); but it is also the most invasive procedure with the highest cost. On the other hand, fecal occult blood test (FOBT) is a low cost test with lower sensitivity. If a FOBT test is positive, then it should be followed up with a colonoscopy to confirm the result. Some screening programs propose to combine these tests to achieve more effective and cheaper results (see e.g. Harper and Jones (2005); Zauber et al. (2008)).

After evaluating alternative programs, the American Cancer Society recommends that average risk people older than 50 are screened with one of the following methods: colonoscopy every 10 years, CT colonography every 5 years, or Double-contrast barium enema every 5 years (American Cancer Society 2016). The idea is to compare the effectiveness and costs of different screening programs that can be composed of a number of tests. The same issues arise when new screening technologies emerge; e.g., Ayer (2015) compares the state-of-the-art screening technology mammography with emerging technologies through a parametric Partially Observable Markov Chain.

In the following we briefly discuss OR-based methods that are used to compare different screening programs.

12.2 OR Models for Evaluation and Optimization of Screening Policies

Screening effectiveness should be evaluated in terms of its effects on the health state of the population, which would be improved by prevention or earlier detection of cancers, and thus result in lower mortality, at the expense of the screening cost. The treatment costs may be lower or higher, depending on the disease. For this purpose, different OR models are developed that capture the disease dynamics and the effects of screening on these dynamics in a population or an individual. One of the points these models share is "health states" although they can be defined at varying levels of detail. In CRC, we can consider the following set of health states: healthy, with polyp, preclinical CRC, clinical CRC and death, as suggested by e.g., Harper and Jones (2005) and Güneş et al. (2015). A more detailed description of health states for CRC is possible, see e.g., Zauber et al. (2008), where states with polyp, preclinical and clinical CRC have additional layers. Another aspect of screening is the target population, and some models include different risk groups, which are incorporated in the health state definitions, e.g., see Erenay et al. (2014) for CRC.

Such models that represent the problem in terms of states and transitions between states are referred to as "State Transition Models" in the medical decision making, health technology assessment and health economics fields (Siebert et al. 2012). These models can be grouped into two according to the level of detail they provide: Individual-based models versus cohort-based models. A detailed comparison of these approaches and best practices on modeling can be found in Siebert et al. (2012), a report that focused on microsimulation and cohort-based simulation models.

Other approaches can be used to develop more general state transition models. For example, Markov Decision Processes (MDPs) are frequently used for personalized medical decision making as they allow optimization of sequential decisions. To evaluate screening programs, general stochastic models are developed. Below we provide a brief description of each approach together with example applications in the literature.

12.2.1 Modeling at the Individual Level

Individual level models, also called microsimulation models, can represent individuals in a population with their distinct characteristics. Therefore, these models tend to be very complex and Monte Carlo simulation or Discrete Event Simulation (DES) are used for their analysis. They typically have three sub-components: Demography, Natural History of the Disease, and Screening. In the first component, a group of individuals are generated, with a date of birth and date of death from causes other than cancer. In the natural history component, there are three steps: (1) Identifying a fixed number of distinct states and characteristics associated with these states; (2) Specifying stochastic rules for transition through states; and (3) Setting values for model parameters (Rutter et al. 2011). Thus each individual's life history is generated in detail so that the time they will get cancer, if ever, is determined, together with the time of death in absence of screening. After the natural trajectories are generated, screening policy can be superimposed on the population and the impact of alternative policies can be evaluated.

In practice, microsimulation models have been widely used to evaluate the costs and benefits of screening. There are several such programs developed by different research groups. A good example for these models is the MISCAN (Microsimulation Screening Analysis) model (Habbema et al. 1985), which has been developed originally for breast cancer and then implemented for colon, cervical, lung and prostate cancer as well. This is a computer program which uses Monte Carlo simulation method to simulate life histories of individuals given the parameters for natural history of the disease. Loeve et al. (1999) explain the MISCAN-Colon model in detail, which has been run with a cohort of 100,000 individuals representing the US population in 1993. The Cancer Intervention and Surveillance Modeling Network (CISNET) is a consortium of researchers using simulation to evaluate interventions for cancer. A comprehensive list of simulation models for different cancers can be seen at the CISNET website (CISNET 2016). Microsimulation models do not allow for interactions between individuals. If that is an important concern as in infectious diseases, then Agent-Based Modeling is used.

There is abundance of examples of microsimulation model implementations in the literature, which are more commonly published in the medical field journals. For a review of microsimulation models see Rutter et al. (2011). These papers commonly develop a simulation of a cohort of individuals to evaluate alternative screening policies using different objectives. For example, Zauber et al. (2008) use microsimulation for colorectal cancer screening with the objective of maximizing the number of life years gained while minimizing the number of colonoscopies over lifetime. They then identify the efficient frontier of screening strategies based on these two criteria.

A Semi-Markov model was developed by Harper and Jones (2005) for CRC, using the standard TNM staging to define the states, which represents the disease stage in terms of the primary tumor size (T), lymph node involvement (N) and information about metastatic deposits (M). This model allows for general

distributions of time for transitions between states. Simulation of a cohort of 1000 individuals is used to evaluate the performance of alternative screening policies with this model, where they estimate a decrease in cause specific mortality by screening in comparison with no screening.

Vieira et al. (2011) develop a microsimulation model for breast cancer screening for 10,000 women, to analyze the length bias in regular interval screening programs. They conclude that the tumors detected by regular interval screening program may be slow growing tumors, while fast growing tumors are missed. Therefore they recommend that the benefits of screening programs should be re-evaluated and future models should account for this length bias.

The models that consider screening schedules usually focus on cancer detection. After a cancer is detected and treated, surveillance of treated patient is done with similar tests as screening, yet with different schedules. Since the disease parameters may be different for recurrent cancers, models should be adapted for surveillance schedules, while similar techniques can be employed. For example Erenay et al. (2016) develop a microsimulation model to develop surveillance strategies for metachronus colorectal cancers (cancers that occur at least after 6 months of treatment). They find a group of pareto-efficient policies and show that the current guidelines can be improved by offering more frequent tests. Zhang et al. (2013) develop a partially observable Markov model for recurrent bladder cancer and compare surveillance strategies via simulations of this model.

Modeling compliance of individuals has become an important concern for evaluation of screening programs in recent years. A common recommendation in these studies is to increase frequency of screening for lower compliant individuals. Brailsford et al. (2012) develop a DES model for breast cancer screening, where they simulate the lives of 1000 women. The model incorporates the compliance behavior of women and shows that accounting for the less-than-perfect attendance to screening may lead to recommendations of different screening policies. For example, they find that increasing attendance rate may be more beneficial than increasing the frequency of screening for breast cancer.

Other examples of microsimulation model applications can be seen at de Kok et al. (2012) and van Ballegooijen et al. (1992) for cervical cancer, Szeto and Devlin (1996) for breast cancer, McMahon et al. (2011) for lung cancer, Gulati et al. (2013) for prostate cancer screening.

Screening can be done for diseases other than cancer, such as diabetes. Diabetes is a chronic disease which increases the risk of Cardiovascular disease (CVD), and may also cause retinopathy at later stages. A review of models on economic evaluation of screening for type-2 Diabetes by Waugh et al. (2007) concludes that there is a strong support for implementing screening for diabetes, since early detection can reduce the risk of CVD. However, more research is needed on the natural history of disease to improve our understanding of benefits of screening and to find optimal screening strategies. Screening for impaired glucose tolerance (IGT), which leads to diabetes if untreated (while if known, life style changes can reduce the risk of diabetes), is also a potentially beneficial strategy. This is similar to CRC,

in the sense that similar to polyps leading to cancer, IGT leads to diabetes and can be treated if detected early, thus preventing diabetes.

There are few studies that model screening for diabetes in the OR literature. Brailsford et al. (2006) develop a DES model for screening for diabetic retinopathy. The model simulates life histories of 100,000 individuals and is embedded within an ant colony optimization to find the optimal screening strategies for diabetic retinopathy. Brailsford et al. (1998) uses an individual based DES model to evaluate screening strategies for early detection of retinopathy in patients with noninsulin dependent diabetes. McLay et al. (2010) use simulation optimization to construct dynamic screening policies for cervical cancer screening.

Hybrid modeling approaches can also be used as in Tejada et al. (2014), which combines the System Dynamics (SD) and DES approaches in a model of breast cancer screening and treatment. In the first phase, DES is used to represent the natural history of the disease for simulated women. In the second phase a combined DES-SD approach is used to represent screening and treatment activities and their impact on the disease progression of individuals, as well as the population level variables that would affect adherence to screening. This framework is similar to a microsimulation approach, where individual life histories are generated and screening policies are imposed on these histories. However their approach also incorporates many variables that would affect adherence such as availability of facilities, distance to facilities, public awareness for breast cancer etc. in the SD model, which feeds into the screening and treatment module. Another hybrid model is developed by Güneş et al. (2004) for breast cancer screening, focusing on the delivery of screening services.

Microsimulation models are advantageous in that they can realistically represent health states in detail, also allowing for history dependency if necessary. However these models tend to be very complex, and developing such models and estimating the corresponding parameters is a challenge.

12.2.2 Modeling at the Cohort Level

Population level (or cohort level) models do not represent distinct individuals in a population. Rather, they keep track of the number of individuals in each state, which we call a compartment. In such models, people flow through states that are relevant for the analysis of the problem. For the screening evaluation problems, typical states are the stages of the natural history of the disease, as well as treatment and death. Typically the transitions depend on a small number of factors such as age, as compared to microsimulation models, which can account for several factors associated with cancer risk.

As a result, there are fewer number of distinct states in a cohort-level model. The population is distributed into those states, assuming these states represent homogenous groups, and they flow throughout these states based on the transition probabilities or deterministic flow rates. At the end of each cycle the population

is re-distributed into these states. These models are Markovian, in the sense that the transitions are not history dependent and in the medical literature, they are commonly referred to as "Markov Models." Allowing for history dependency requires adding new states, which may quickly lead to state explosion. The transitions between states can be deterministic, as in System Dynamic Models, or stochastic, as in stochastic compartmental models. Different methods such as Monte Carlo simulation, or Discrete Event Simulation can also be used to analyze such models.

Güneş et al. (2015) analyze CRC cancer progression when colonoscopy services have limited capacity, through compartmental models. They first consider a simple deterministic model with four compartments as shown in Fig. 12.1, which is extracted from Güneş et al. (2015).

These compartments categorize the population with respect to two health states ("preventable" and "with cancer") and two service system states ("waiting for service" and "after service"). "Preventable" represents the low-cost patients who are either in a pre-cancer or early-cancer stage. "With cancer" represents high-cost patients who are in late-cancer stage. The transition rates are given on the arrows connecting the boxes. Note that the "Preventable" state includes the polyp state, which can be removed by colonoscopy to prevent cancer. Such population dynamics models can be represented as a set of differential equations governing the flows between the states. We denote the number of people in each compartment at time t by $X_i(t)$. The population dynamics in this model are represented by the following differential equations:

$$\frac{dX_1(t)}{dt} = \delta - (\theta + \omega_0 + c\mu)X_1(t) \tag{12.1}$$

$$\frac{dX_2(t)}{dt} = \theta X_1(t) - (\omega_0 + \omega_3 + (1-c)\mu)X_2(t) \tag{12.2}$$

$$\frac{dX_3(t)}{dt} = c\mu X_1(t) - (\omega_0 + \omega_1)X_3(t) \tag{12.3}$$

$$\frac{dX_4(t)}{dt} = (1-c)\mu X_2(t) - (\omega_0 + \omega_2)X_4(t) \tag{12.4}$$

The equilibrium of this system can be found by setting $\frac{dX_1(t)}{dt} = \frac{dX_2(t)}{dt} = \frac{dX_3(t)}{dt} = \frac{dX_4(t)}{dt} = 0$. The number of people in each compartment in equilibrium is then found by solving the above system of equations. Based on this, performance measures such as number of cancer deaths can be obtained and compared under different screening policies, possibly represented with different transition rates between compartments.

This example is quite simple, although unrealistic, and hence a closed-form solution can be obtained. Analysis of such models can generate insights concerning the optimal screening policies. For example, the analysis of the model in Fig. 12.1 shows that it is optimal to allocate some of the resources to screening services, when the screening test prevents cancer for good. Although this condition is an extreme

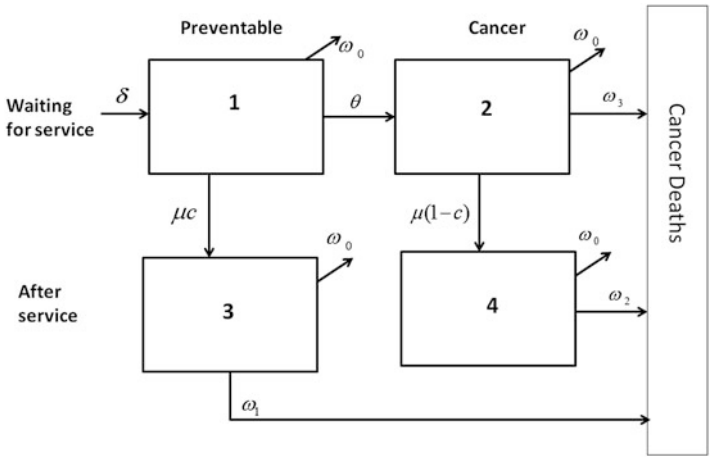

Fig. 12.1 A simple compartmental model for prevention/diagnosis dynamics for CRC. Figure is taken from Güneş et al. (2015)

case, the result offers an understanding of the dynamics that lead to more or less allocation of resources for screening.

In more realistic models, the number of states increases and usually obtaining closed-form solutions is not possible. In those cases, either numerical integration, or a simulation of the system yields a solution for the behavior or the system. Güneş et al. (2015) provide an extended model with 23 compartments, as shown in Fig. 12.2. The states are presented as boxes on a matrix with the health state on the horizontal dimension and the status in service system on the vertical dimension. The transition rates are given on the arrows connecting the boxes. This model assumes four health states, distinguishing between healthy, polyp, early stage cancer and late stage cancer, and uses cancer mortality as the performance measure to evaluate resource allocation policies. Screening process aims to detect and remove polyps so as to decrease the number of cancers, the rate of which depends on the test sensitivity and resource capacity as well as the compliance rate. This model includes a waiting stage for colonoscopy, during which the health state can deteriorate as shown by the horizontal transitions, thus captures the effect of insufficient resources on health state dynamics.

David M. Eddy is the pioneer for using such mathematical models for evaluation of screening programs, and has won the INFORMS Lanchester Prize in 1980 for his book *Screening for Cancer:Theory, Analysis and Design* (Eddy 1980). He has developed mathematical models for breast cancer, colorectal cancer, cervical cancer, lung cancer. These models are mostly Markovian models and are described as a set of differential equations that represent the disease trajectories, which are then solved by numerical integration. The generic model for timing repeated tests is presented in Eddy (1983), which is the basis for most of his latter applications on cancer screening. Eddy (1990) explains the results of such a mathematical model and its use

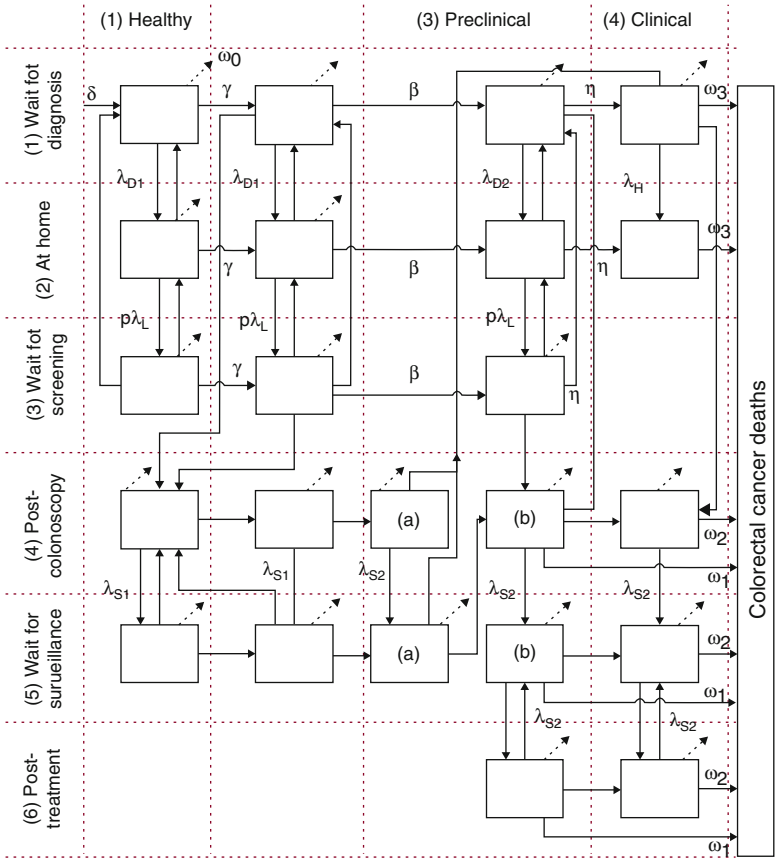

Fig. 12.2 A population dynamics model for CRC when the capacity for colonoscopy services are limited. Figure is taken from Güneş et al. (2015)

for evaluating screening strategies with different tests for CRC. The model consists of a nine-state time-varying Markov chain with health states of individuals defined as: alive and without cancer, have cancer diagnosed in various stages, have died from cancer, and died from other causes.

Frazier et al. (2000) develop a Markov model for CRC screening to estimate cost effectiveness ratios for alternative screening policies. The model simulates the evolution of health states from healthy to adenomatous polyp to cancer. A screening mechanism is superimposed on the natural history of colorectal disease. Individuals representative of the 50-year-old US population, were placed into health states (low risk, high risk, localized, regional, or distant cancer) and they flow through these states on an annual basis based on transition probabilities estimated from the literature.

Macafee et al. (2008) develop a 19 state Markov model for CRC screening with FOBT to assess the impact of an aging population on cost effectiveness of screening. Simulating the screening and treatment process for two cohorts representing the year 2003 and 2033, where the latter is an older cohort, they find that cost-effectiveness of CRC screening may improve with increasing life-expectancies. Siebert et al. (2006), developing a 16-state Markov model for cervical cancer screening provide another example for cohort-based models.

Such models can also be used for modeling screening for infectious diseases. For example, Brandeau et al. (1993) develop a dynamic compartmental model formulated as a set of simultaneous nonlinear differential equations, and solve it numerically to determine screening policies for HIV for women of childbearing age. They conclude that screening of the low-risk women is not beneficial, while it may benefit the medium to high risk groups. There is an extensive literature on infectious diseases, which is not our focus in this chapter. We refer the interested readers to Brandeau (2013).

Jones et al. (2006) provide an example for the use of System Dynamics Simulation that may help to understand the diabetes population dynamics at a conceptual level. The model illustrates that there may be an unintended increases in diabetes prevalence due to diabetes control, and the diabetes control efforts alone are not able to reduce diabetes-related deaths in the long term. More research is needed to explore such dynamics.

Another interesting issue is how the screening policies are implemented by the physicians in practice. Karanfil (2016) addresses the discrepancies between guidelines and the actual practice in the case of prostate cancer. After observing oscillations in decision thresholds of clinical guidelines, and persistent over-under screening through an empirical study, she examines the relations between the evidence-based guidelines and actual practice using system dynamics and simulation models.

Compartmental models are easier to develop and use compared to microsimulation models. However, the limited number of states limit the realism of these models, and it may be a challenge to determine the set of states that is a good enough representation of reality. Estimation of parameters is also a critical component of this type of modeling.

12.2.3 Stochastic Models for Screening Program Evaluation

Stochastic models for evaluating screening programs have appeared in the OR literature since the 1960s. The earliest study we know of is Lincoln and Weiss (1964) who propose a stochastic model to evaluate the effects of recurrent medical examinations with imperfect results. They illustrate the model results by contrasting them with an incidence curve of cervical cancer. Zelen and Feinleib (1969) develop a stochastic model to evaluate the effects of the lead time gained by detecting a disease in a pre-clinical health state through a screening programme. Most of

the early work focus on breast cancer: Shwartz (1978) develop a mathematical model that represents the disease progression to evaluate the effects of different screening policies with respect to various benefit measures such as life expectancy or probability of no recurrence. Özekici and Pliska (1991) develop a so-called delayed MDP model to minimize total expected cost which is calculated as the sum of inspection, false positive, treatment and mortality costs. Baker (1998) parametrically represents the processes of tumor origination and growth, detection of tumors at screening, treatment and survival of women with cancers after diagnosis. Other early studies on breast cancer include Voelker and Pierskalla (1980) and Parmigiani (1996). Finally, Lee and Pierskalla (1988) analyze the test choice and screening periods for contagious diseases.

There is a stream of literature that applies dynamic programming, or heuristic-based solution approaches for screening program evaluation. For example, Yang et al. (2013) develop a dynamic programming model to determine the screening policy and thresholds for treatment assuming only three screenings, for childhood obesity. Unlike most dynamic programming applications that focus on an individual, this study uses the population-wide distribution of the disease as a state definition. They optimize over the class of biennial threshold policies, which measures every child's Body Mass Index (BMI) every 2 years and treats if the (BMI exceeds a threshold. To this end, they model the Markovian evolution of the probability density function of the BMI in the population of children.

Rauner et al. (2010) propose a mathematical model of chronic disease policy that accounts for population health dynamics as well as program costs. They use a metaheuristic approach applying Pareto Ant Colony Optimization algorithm to find pareto-optimal screening schedules for risk groups by considering cost and effectiveness outcomes as well as budget constraints. This approach is illustrated with a numerical example for breast cancer, where cost-effective screening strategies are presented for different budget constraints.

Lee et al. (2015) evaluate screening policies for hepatocellular carcinoma (HCC), the most common form of liver cancer in adults. They develop reinforcement learning techniques that learn about the biomedical dynamics of the patient to be able to correctly assess the HCC risk. Three different screening policies are proposed and evaluated when the screening capacity is limited. Chen et al. (2016) on the other hand analyze a set of surveillance policies for HCC in a mixed-integer programming (MIP)-based framework and find the surveillance policies with the maximum societal net benefit. Their findings show that the surveillance policies should depend on the stage of hepatitis C infection and age.

12.2.4 OR Models for Medical Decision Making

OR models have been used effectively to support medical decision making for the last few decades. In these models, an individual moves through a number of health states that can be affected by screening programs. Typically, the number of these health states is less than the OR models reviewed in the previous sections.

The key in developing good OR models in this field is to represent the disease progression effectively under different screening decisions. The objective of these models can be to minimize total expected cost, including screening and treatment costs among others, or to maximize the total expected QALY, or a combination of both. For example, Zhang et al. (2012b), who optimize prostate-specific antigen (PSA) screening decisions, use two objectives: (1) to maximize the total expected QALYs, which represents the patient perspective, and (2) to maximize the expected monetary value based on societal willingness to pay for QALYs and the cost of PSA testing, prostate biopsies, and treatment, which represents the societal perspective. One of the major challenges in this type of modeling is the estimation of model parameters, as it is essential to represent the disease progression correctly, especially to draw right insights and conclusions. Hence, a thorough analysis of available health statistics and/or a close contact with medical doctors are required to develop and validate these models.

Alagoz et al. (2011) present a review of the literature on cancer screening, whereas Alagoz (2014) focus on optimizing screening decisions through POMDP models. More recently, Steimle and Denton (2016) review the use of MDP models for improving screening and treatment of chronic diseases. We will first review relatively recent studies on a variety of diseases, then focus on the CRC screening, where we provide both a review and a simple POMDP model for illustration purposes.

Breast cancer screening is one of the most commonly studied areas in medical decision making. Ivy (2009) provides a review of OR models for breast cancer screening decision problem. Here, we review more recent studies. Maillart et al. (2008) develop a partially-observed Markov chain model to evaluate different mammography screening strategies for premenopausal versus postmenopausal women, with respect to the survival probability and the expected mammogram count. In Chhatwal et al. (2010), the problem of requesting a mammogram versus a biopsy is considered, where an MDP model is developed to aid radiologists in their biopsy decision-making process. Later on, this model is extended to account for budgetary constraints by Ayvaci et al. (2012). Ayer et al. (2012) formulate a POMDP model to individualize mammography screening decisions for various risk groups with the objective of maximizing total expected QALYs. Lately, there are studies on individual adherence behavior and its effects on breast cancer screening, e.g., Madadi et al. (2015) and Ayer et al. (2016). Kirkizlar et al. (2010) also model patient behavior with an MDP model for Hepatitis C screening and treatment, which incorporates the health state as well as the belief of patient about this health state. It is assumed that this belief can have an impact on the future behavior of the patient, thus affect the disease progression.

Prostate cancer is another cancer type analyzed by OR models. Zhang et al. (2012a) build a POMDP model to optimize the prostate biopsy referral decisions with the objective of maximizing the total expected QALYs, where these decisions depend on the PSA test results. The interested readers can find a review on the OR models analyzing prostate cancer screening, detection and treatment in the recent review article Price et al. (2016).

Finally, Onen et al. (2016) present a non-cancer example in medical decision making: they build an MDP model with an objective combining quality-adjusted life years and costs, to investigate optimal population screening policies for Alzheimer's disease. Although screening is not recommended in the base case which represents the current practice, it is found that the optimal policy is very sensitive to treatment effectiveness. Hence, when "better" treatment plans are developed, implementing a population screening policy may be socially optimal.

In the context of colorectal cancer screening, only partially observable models are used: Leshno et al. (2003) build a POMDP model to analyze the cost-effectiveness of various screening strategies, and conclude that annual FOBT and sigmoidoscopy during a 5-year interval as well as one-time colonoscopy screening are cost-effective. Li et al. (2014) compare the performances of eight different screening policies, differentiated by screening frequency, initial and end screening ages, as well as screening compliance rate, through a partially-observed Markov chain model. Erenay et al. (2014) use a POMDP model to develop optimal CRC screening decisions with colonoscopy for patients having different risk factors, genders and health histories. They propose a more aggressive screening policy for certain individuals, compared to the one suggested by the guidelines. Moreover, their recommendations show that females without CRC history should be screened less frequently than males, whereas females with CRC history should be screened more frequently than males. Finally, Li et al. (2015) build a POMDP model that accounts for individual adherence behavior to evaluate the following screening strategy: annual FOBT for the average-risk patients and colonoscopy for the high-risk and the average-risk patients with positive FOBT results. Their results suggest a more aggressive colonoscopy screening policy for those with low compliance rate.

A POMDP model can be posed as a finite-horizon or an infinite-horizon problem. Due to the longer time intervals between decision epochs and finite lifetime of humans, finite-horizon problems are more common when screening decisions are considered. At each decision epoch, which can be an absolute time such as years or a relative time such as the time of obtaining a certain test result, the decision maker has to choose one of available actions using all the available information. A POMDP model consists of states, actions, observations, belief states, and rewards.

Now we present a simplified version of the POMDP model in Erenay et al. (2014). We assume that there is only one of type of patients, so there are no risk groups or patients who recover from CRC. Accordingly, we maximize the total expected QALYs of a patient until he starts CRC treatment or dies, where we assume that he will live for at most 100 years so that the problem is defined over 50 years after age 50. Here are the details of the POMDP components:

States Figure 12.3 presents the health states of a patient and the transitions within them with respect to CRC. The states of the POMDP model are H (healthy), P (with polyp), C (CRC), T (CRC treatment) and D (death). The natural CRC progression starts with a healthy person developing a polyp, which happens with probability ρ_{HP}^t in year t. Then, the polyp(s) may grow into a tumor with probability ρ_{PC}^t. These parameters describe the disease progression, when there is no interference,

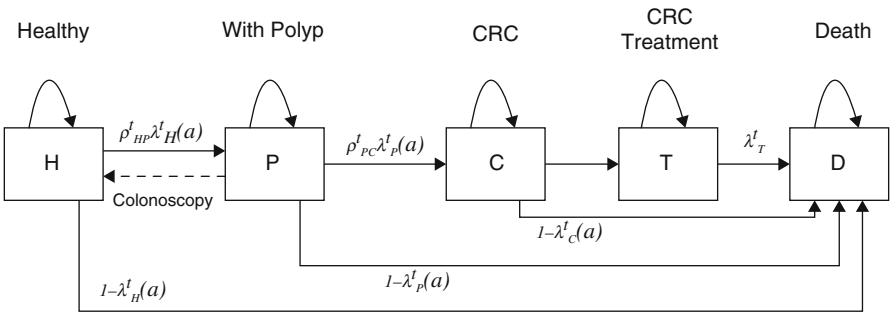

Fig. 12.3 Health states in CRC

where they are depicted as regular arcs in Fig. 12.3. Note that the transition from C to T occurs upon detecting the CRC, which depends on the decisions in a more complicated way so that Fig. 12.3 does not indicate a rate for this arc. Colonoscopy screening not only detects a CRC at an early stage, but it also prevents it by removing polyps before they become cancerous. The dotted line from state P (with polyp) to state H (healthy) indicates this possibility. This is the only time that the CRC progression can be reversed. If not detected, polyps can turn into tumors, which can be diagnosed either due to cancerous symptoms or by screening. When diagnosed, the patient undergoes a CRC treatment, while it is also possible for the patient to move to the death state before any diagnosis. The death state is an absorbing state, that is reachable from all other health states. The survival probability depends not only on the health state (s) but also on the action taken (a) and year t, denoted by $\lambda^t_s(a)$. Hence, $1-\lambda^t_s(a)$ is the probability that the patient dies within a year when he starts in the state-action pair (s,a) in year t.

The set of all states is denoted by $\mathscr{S} = \{H, P, C, T, D\}$. An individual who has no symptoms can be healthy (H), with polyp (P) or with CRC (C), so that they are indistinguishable from each other. On the other hand, states T and D are completely observable. We denote the set of partially observable states by $\mathscr{P} = \{H, P, C\}$, and the set of completely observable states by $\mathscr{C} = \{T, D\}$.

Order of Events, Actions and Observations Year t is defined as the t^{th} year after age 50 with $0 \le t \le 50$. The POMDP model assumes the following order of events in year t: A patient decides whether to undergo colonoscopy (c) or not (n) at the beginning of year t (so that the beginning of each year is a decision epoch). We let $\mathscr{A} = \{c, n\}$ be the action set. If a patient with CRC chooses to do nothing, then he either finds out about his CRC himself by developing CRC symptoms (denoted by $++$), or does not (denoted by $-$). A patient who does not have CRC will always have a signal of $-$ under action n. Note that polyp(s) does not cause any symptoms, so a patient cannot detect polyp(s) on his own.

If the patient chooses to undergo colonoscopy, there are four possible outcomes:

- Colonoscopy does not find any polyp or CRC (denoted by $-$), which may happen in states H, P and C.
- It finds and removes polyp(s) (denoted by R), which may happen only in state P.
- It detects a CRC (denoted by $+$), which may happen only in state C.
- The patient detects a CRC himself ($++$), which may happen only in state C.

We define $\mathscr{O} = \{++, +, R, -\}$ as the set of all possible outcomes. After the outcome of the chosen action is observed, the year passes during which the health state changes according to the disease progression parameters. At the end of year t, the belief states for year $t+1$ are updated according to the belief state, action and observation in year t.

Belief States If the state in year t is not observable, we only have the information about the history, i.e., an initial probability distribution about the health state, as well as the actions and observations until year t. However, the history can be summarized by a probability distribution on $\mathscr{P}$, which is shown to be a sufficient statistic by Smallwood and Sondik (1973). Accordingly, we define a belief state of the POMDP model as the probability that the individual is in state H, P or C in year t, denoted by $\bar{b}^t = (b^t_H, b^t_P, b^t_C)$. The definition of belief states transforms a POMDP model to a continuous-state MDP model.

In our simplified model, whenever the state becomes observable, it remains observable, and there are no further screening decisions. Hence, the process becomes a regular discrete-time Markov chain in $\mathscr{C}$.

Transitions Between Belief States The derivation of the transition probabilities between the belief states requires calculating certain probabilities. More explicitly, we need to compute the following quantities, when the year starts with the state-action pair (s, a), where $s \in \mathscr{P}$ and $a \in \mathscr{A}$:

1. The probability of observing each outcome o ($f^t(o|s,a)$): Let ω denote the probability that a patient discovers his own CRC. Then for action n, we have $f^t(++|C,n) = \omega, f^t(-|C,n) = 1-\omega$, and $f^t(-|s,n) = 1$ for $s \in \{H, P\}$.

   The computation of $f^t(o|s,c)$ uses the sensitivity of colonoscopy for polyps and CRC, denoted by α_P and α_C, respectively. If a patient in state P (C) undergoes colonoscopy, his polyp(s) (CRC) will be detected with probability α_P (α_C), resulting in observation R ($+$), i.e., $f^t(R|P,c) = \alpha_P$ and $f^t(+|C,c) = \alpha_C$. For state P, the only other possible outcome is $-$. In state C, both self-detection and $-$ are possible. We assume that colonoscopy screening takes place in the beginning of the year, so that self-detection is possible only if the outcome of the colonoscopy is negative; accordingly $f^t(++|C,c) = (1-\alpha_C)\omega$. Table 12.2 in Appendix presents the values of $f^t(o|s,a)$.
2. The transition probability to state s' when outcome o is observed ($p^t(s'|s,a,o)$): The transitions from a state s to state s' depends on both actions and observations. For example, consider a patient with polyp: when the observation is $-$, then s/he

will stay in P with probability $\lambda_P^t(a)(1-\rho_{PC}^t)$, i.e., $p^t(P|P,a,-) = \lambda_P^t(a)(1-\rho_{PC}^t)$ for both $a = n, c$. However, if action c is chosen and outcome R is observed, the polyp(s) are removed moving the patient to healthy state, so that $p^t(P|P,c,R) = \lambda_P^t(c)\rho_{HP}^t$. The other probabilities can be found in a similar way. Table 12.3 in Appendix tabulates the values of $p^t(s'|s,a,o)$.

3. The transition probability to state s' ($g^t(s'|s,a)$): Once the probabilities $f^t(o|s,a)$ and $p^t(s'|s,a,o)$ are known, it is straightforward to calculate $g^t(s'|s,a)$ by considering all observations that can be obtained in state-action pair (s,a):

$$g^t(s'|s,a) = \sum_{o\in\mathscr{O}} p^t(s'|s,a,o) f^t(o|s,a).$$

The resulting values are given in Tables 12.4 and 12.5 in Appendix for actions n and c, respectively.

Now we can derive the transition probability of moving from a belief state $\bar{b}^t$ to belief state $\bar{b}^{t+1}$, where $b_{s'}^{t+1} = b'(s'|\bar{b}^t,a,o)$ is the probability that the next state will be $s' \in \mathscr{P}$ given that the belief state is $\bar{b}^t$, action a is taken and outcome o is observed at time t. In what follows, only the transition from year t to $t+1$ will be considered, and the belief states $\bar{b}^t$ and $\bar{b}^{t+1}$ will be denoted by $\bar{b}$ and $\bar{b}'$, respectively.

A belief state is essentially a probability distribution defined over all partially observable states, hence excluding the probability of moving to the observable states. However, there is always a positive probability that the health state moves to T or D. Therefore, we define $\tilde{b}(s'|\bar{b},a,o)$ as the conditional probability that the state in year $t+1$ will be $s' \in \mathscr{S}$ given that the belief state is $\bar{b}$, action is a and outcome o is observed in year t. We will use $\tilde{b}$ to denote the vector $(\tilde{b}(s|\bar{b},a,o) : s \in \mathscr{S})$.

We first compute $\tilde{b}(s'|\bar{b},a,o)$:

$$\tilde{b}(s'|\bar{b},a,o) = \frac{\sum_{s\in\mathscr{P}} p^t(s'|s,a,o) f^t(o|s,a) b_s}{\sum_{s\in\mathscr{P}} f^t(o|s,a) b_s}, \text{ for all } s' \in \mathscr{S}. \tag{12.5}$$

The denominator of (12.5) is the probability of observing outcome o when the belief state is $\bar{b}$ and action a is taken in year t. The numerator, on the other hand, corresponds to the probability that the state is s' in year $t+1$ and outcome o is observed in year t, when the belief state is $\bar{b}$ and action a is taken in year t.

To calculate the updated belief state, $b'(s'|\bar{b},a,o)$, we need to condition on the event that the state remains in $\mathscr{P}$ in year $t+1$. Denoting the probability of this event by $\ell(\bar{b},a,o) = 1 - \tilde{b}(T|\bar{b},a,o) - \tilde{b}(D|\bar{b},a,o)$, the updated belief state is:

$$b'(s'|\bar{b},a,o) = \frac{\tilde{b}(s'|\bar{b},a,o)}{\ell(\bar{b},a,o)}, \quad \text{for all } s' \in \mathscr{P}. \tag{12.6}$$

Rewards Our objective is to maximize the expected total QALY (TQALY). Accordingly, the immediate reward in year t is set to the difference between the expected lifetime accrued and the expected disutility measured in life years incurred

in year t. This reward depends on the current (s) and the next states (s'), the observation (o), the action taken (a) as well as the year (t), which is denoted by $q^t(s,a,o,s')$. The calculation of $q^t(s,a,o,s')$ requires a thorough analysis of CRC statistics. We assume that we already have these parameters, and refer the interested readers to Erenay et al. (2014).

The POMDP model accounts for the QALYs for each state s and action a, so we calculate the expected QALY for all pairs (s,a) in each year t as follows:

$$q^t(s,a) = \sum_{s' \in \mathscr{S}} \sum_{o \in \mathscr{O}} p^t(s'|s,a,o) f^t(o|s,a) q^t(s,a,o,s'), \text{ for all } s \in \mathscr{P}.$$

In state D, there is no remaining QALY. In the other observable state T, there are no actions and no possibility to move to the partially observable states. Hence, the expected discounted TQALY in state T, denoted by $Q^t(T)$, is constant and known.

Optimality Equations We define the optimal value function, $V^t(\bar{b})$ as the maximal expected TQALYs over the years between t and $t+50$ until the patient starts CRC treatment or dies when the belief state in year t is $\bar{b}$. Then the optimality equations are given as follows:

$$V^t(\bar{b}) = \max_{a \in \mathscr{A}} \left\{ \sum_{s \in \mathscr{P}} b_s q^t(s,a) + \right.$$
$$\left. \beta \sum_{s \in \mathscr{P}} \sum_{o \in \mathscr{O}} b_s f^t(o|s,a) \left(\ell(\bar{b},a,o) V^{t+1}(\bar{b}') + \tilde{b}(T|\bar{b},a,o) Q^{t+1}(T) \right) \right\},$$

where β is a discount factor. The first term is the expected QALY accrued in year t, whereas the second term refers to the future QALYs. The summation considers the probability of being in state s, and of observing outcome o given the state-action pair (s,a) for all $s \in \mathscr{P}$ and $o \in \mathscr{O}$. Given that outcome o is observed in state s with action a, the state may either remain in set $\mathscr{P}$, which happens with probability $\ell(\bar{b},a,o)$, or move to set $\mathscr{C}$ with probability $1-\ell(\bar{b},a,o)$. In the former case, the belief state becomes $\bar{b}'$ in year $t+1$, due to (12.6). In the latter case, the state becomes either D, with probability $\tilde{b}(D|\bar{b},a,o)$, incurring no further rewards, or T, with probability $\tilde{b}(T|\bar{b},a,o)$, incurring a fixed reward of $Q^{t+1}(T)$.

The solution methods for POMDPs are reviewed by Lovejoy (1991). When $V^t(\bar{b})$'s are piecewise linear and convex (PW&C), the optimality equations can be solved by Monahan's algorithm (Monahan 1982) with Eagle's reduction or one of its variants (Eagle 1984). Our POMDP model is a special case of the model in Erenay et al. (2014), who have shown that the resulting value functions are PW&C. Hence, our POMDP model can be solved with one of the above methods.

This POMDP, like the other POMDP models in medical decision making, has many parameters to be estimated from the available data and/or with expert opinion. The model needs to be validated (and calibrated, if necessary) by comparing the

model results with those from microsimulation models and through expert opinions. After validation, the POMDP model is used to construct a thorough numerical study which compares the optimal policies with the guidelines and makes further recommendations.

12.3 OR Models for Management of Screening Services

In the previous section, we reviewed the models focusing on the medical decisions at the population level or individual level. After a screening program is evaluated and found cost effective, resources should be deployed to deliver screening to the target population. If a test is found cost-effective for a target group of population, it is generally recommended by health societies, and eventually included in insurance coverage packages. A screening program may be set as a priority by the public health authorities, with the aim of increasing the percentage of screened population. For example there are organized screening programs for colorectal cancer in many developed countries.

Increasing effectiveness of screening programs is a challenge for many reasons. Availability and accessibility of the screening test is one of the most important factors that affect participation and program success. Providing high quality tests is another important goal that would help achieving the screening program objectives. While all of these objectives are important, there may also be trade-offs between them, as demonstrated in Güneş et al. (2004). That paper presents a hybrid model of SD and DES, applied to breast cancer screening, focusing on management of screening processes. Specifically, the trade-offs between outreach, quality, and congestion at the clinics are modeled and numerical experiments highlight the importance of ensuring high test quality and providing sufficient service capacity to minimize mortality. Therefore, it is evident that effective planning and management of screening services is essential for achieving screening objectives.

Service delivery management related decisions include allocation of funds to screening, location of facilities, allocation of staff and equipment to these facilities, allocation of residents to facilities, allocation of the available resources to screening demand, and patient scheduling. In this section we provide a review of papers applying OR to these problems in the screening services contexts. Most of the papers can be grouped under two categories: location-allocation problems, and resource allocation problems. In the remaining, we briefly review recent research under these headings.

12.3.1 Location Models for Screening Facilities

OR techniques have been applied successfully for location of healthcare facilities. Güneş and Nickel (2015) provide an overview of location problems in healthcare.

For a recent review of location models for preventive healthcare, see Verter and Zhang (2015). Here we briefly summarize the fundamental aspects of these models that are specifically applicable to location of screening facilities.

There are two distinguishing features of location problems for screening services: (1) Demand is modeled as a function of distance and/or time to get service, and (2) Minimum volume requirements may be imposed to allow opening facilities as a quality appraisal requirement.

People demanding screening tests are not in an urgent need, therefore providing their participation to screening remains a challenge. It has been shown that the accessibility of facilities influence willingness to get the screening test (Zimmerman 1997). Therefore, accessibility of the service facilities is an important determinant of screening program success. Location models for screening services usually incorporate a demand function decaying with distance (Verter and Lapierre 2002; Zhang et al. 2012), or distance plus waiting at the clinic (Zhang et al. 2009, 2010).

The second important distinguishing feature of screening services is the relationship between quality of the test and the number of tests performed at a clinic. This volume-quality relationship is well-known for other healthcare services as well, such as surgeries. For breast cancer screening, a radiologist should interpret at least a given number of mammograms every year to have accreditation (this number varies, for example, 960 for the US (Food and Drug Administration 2011), 5000 for the UK (Wilson and Liston 2011)). If relevant, such a minimum workload requirement should be included in the optimization models, as in Verter and Lapierre (2002).

Here we present the model by Verter and Lapierre (2002), which is a model for preventive healthcare facility location, applied to location of breast cancer screening facilities in Montreal, Canada, and location of public health centers in Atlanta, US. Let I be the set of population centers and J denote the set of alternative facility sites. The participation function c_{ij} defines the expected number of residents from population center i that would seek service at facility site j. The model assumes that participation function c_{ij} is a linear decaying function of distance, where participation is zero beyond travel distance of D units, D representing the maximum distance that an individual would travel for screening services. For each pair ij, the set S_{ij} defines the set of alternative facility cites, denoted by l, which are closer to population center i than site j is.

$$\text{Maximize} \quad \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} \tag{12.7}$$

$$\text{subject to} \quad \sum_{j \in J} x_{ij} \leq 1 \qquad \forall i \in I \tag{12.8}$$

$$\sum_{i \in I} c_{ij} x_{ij} \geq W_{min} y_j \qquad \forall j \in J \tag{12.9}$$

$$x_{ij} \leq y_j \qquad \forall i \in I, j \in J \tag{12.10}$$

$$x_{ij} \leq 1 - y_l \qquad \forall l \in S_{ij} i \in I, j \in J \tag{12.11}$$

$$y_i \in \{0,1\} \qquad \forall i \in I \tag{12.12}$$

$$x_{ij} \in \{0,1\} \qquad \forall i \in I, j \in J, \tag{12.13}$$

Here the decision variables are: $y_j = 1$ if a facility is open at site j, 0 otherwise, and $x_{ij} = 1$ if population center i is served by a facility at site j, 0 otherwise. The objective function maximizes the expected number of people participating to screening program. Constraints (12.8) ensure that each population center is assigned to at most one population center, while letting a population center remain unassigned if the closest open center is farther than D units. Constraints (12.9) ensure that the open facilities are assigned at least W_{min} people, to satisfy the minimum workload requirement. Constraints (12.10) ensure that a population center is served by an open facility and Constraints (12.11) ensure assignment to the closest open facility. Finally the integrality requirements are given by (12.12) and (12.13).

Other examples of location models for screening facilities can be seen in Davari et al. (2016), proposing an efficient variable neighborhood search procedure to solve the preventive healthcare facility location problem with budget constraints and Akhundov (2015), developing a simulation-based optimization framework for the location and staffing decisions for screening facilities. Finally, Gu et al. (2010) define a new accessibility measure for the preventive healthcare facility location problem and solve it using an interchange algorithm.

12.3.2 Resource Allocation for Screening Services

Resource allocation problems for screening services arise at different levels: at the strategic level the investment decision is made where a common budget is to be allocated to screening services and other services. At the tactical and the operational levels, scarce resources such as staff and machinery are allocated to screening vs. other services.

Frequently, provision of screening tests uses the same resources with other healthcare services. For example, colonoscopies and mammographies can be used for both screening and diagnostic purposes for colorectal cancer and breast cancer respectively. Physicians interpreting the test results provide other services as well. As a result, a rationing of resources for screening services is needed at the tactical and operational level. In the literature, there are few OR applications for the resource allocation problem for screening services.

Güneş et al. (2015) develop a compartmental model which accounts for the disease progression dynamics and a shared resource for screening and diagnosis of colorectal cancer, and test alternative static capacity allocation schemes. Their model is briefly explained in Sect. 12.2 and illustrated in Fig. 12.2. They show via numerical experiments that it should be ensured that sufficient resources are allocated for diagnosis of CRC, while the remaining resources should be used for screening.

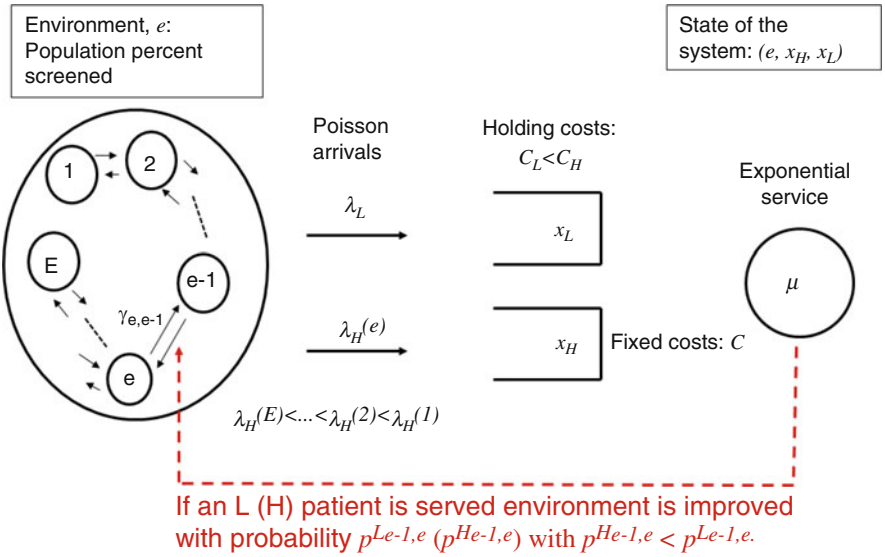

Fig. 12.4 Illustration of the MDP model for colonoscopy services

In Örmeci et al. (2015), optimal dynamic scheduling of a resource that is shared between screening and diagnosis services is modeled using Markov Decision Processes. This stylistic model uses an environment variable to represent the effects of scheduling decisions. The "environment" is a dynamic variable which can be influenced by providing more or less screening services, and it determines the risk of disease in the population, which in turn influences the demand for diagnostic services. Figure 12.4 illustrates the MDP model for colonoscopy scheduling, where the environment state (e) is interpreted as the percent of population screened. The state of the model also includes the number of patients waiting for screening (x_L) and the number of patients waiting for diagnosis (x_H). The environment state (e) worsens with a rate of $\gamma_{e,e-1}$, while it can be improved by providing a screening (diagnostic) colonoscopy with probability $p^L_{e-1,e}$ ($p^H_{e-1,e}$), where $p^H_{e-1,e} < p^L_{e-1,e}$. The system incurs a fixed cost of C for each patient who requests a diagnostic colonoscopy to cover the expected treatment costs, as well as a holding cost per unit time per patient (c_L for screening and c_H for diagnosis). Implementation of this framework for colonoscopy scheduling problem for CRC screening leads to the conclusion that high risk individuals who ask for a diagnostic colonoscopy (possibly as a result of a positive result from another test) should be given priority over individuals asking for a regular screening colonoscopy.

A stream of literature analyzes the resource allocation problem in the context of allocating funds to preventive services and or treatment of diseases. Majority of the papers in this stream are for infectious diseases, and they use population dynamics models to represent disease dynamics. Effect of investments on health outcomes are evaluated using the disease models. Although these models are generally not for

screening services, since they are general models, a similar framework could apply to screening services as well. We therefore review a few examples from this stream here.

Alistar et al. (2013) develop a mathematical model for the decision of allocating investments for treatment and prevention of HIV. An epidemic model first defines the disease dynamics using differential equations. Then the parameters in this epidemic model are modeled as functions of investments, and an optimization model is solved to characterize optimal investment policies. Brandeau and Zaric (2009) develop a framework to asses the optimal investment rate for a prevention program. Using a population dynamics model for HIV prevention and the sufficient contact rate as an outcome of this model, they analyze different production functions for investment, which model the return on investment as the change in the sufficient contact rate. They show that a prevention program being cost-effective does not mean that all of the budget should be spent on that program; when the marginal cost per infection averted is increasing (or equivalently, returns on investment is decreasing), more investment is not always better. Malvankar-Mehta and Xie (2012) develop an incentive-based resource allocation model for the prevention of HIV in multiple populations. Using a game-theoretic framework, they model the fund allocation problem at each decision making level as an optimization problem that considers the incentive mechanisms.

12.4 Conclusion

Screening for chronic diseases is an important public health service. Many decisions are given at different levels of planning this service. OR tools have proven to be very useful to help decision making for screening services. This chapter provided an overview of applications in different aspects of planning and delivering screening services.

Most of the reviewed studies focused on medical aspects of the screening services. At the public policy level, cost effectiveness of screening for different diseases has attracted much attention. At the individual medical decision making level, personalized screening schedules have been developed. Understanding the disease progression is one of the key factors that facilitate developing OR models, where the knowledge on the dynamics that govern how a disease evolves and the ability to estimate the corresponding parameters are essential. Moreover, for an OR model to be useful there should be trade-offs concerning the screening decisions, such as costly screening tests, risk of over-diagnosis, or uncertain benefits of early detection. We can observe that most screening problems that attracted attention from the OR community have such trade-offs involved, such as cancers of the breast, colon and prostate.

Other diseases which are included in recommendations for screening such as lung cancer, type 2 diabetes and hypertension are not widely studied by the OR community. One common property of these diseases is the fact that their

risk can be significantly decreased by lifestyle changes, which are affected by behavioral factors that are not fully understood yet, making it difficult to include in OR models. For lung cancer, the existing screening tests have significant risks due to additional radiation and a high number of false positives that call for more scans and (possibly invasive) tests. Technological advances in the screening methods (such as discovering new biomarkers for lung cancer) may make studying this disease more attractive for the OR community. Open research questions for diabetes screening include finding the optimal screening intervals, choosing the right screening methods and investigating resource requirements to carry out population screening programs.

Future research directions could include modeling such understudied diseases. Another understudied area is the recurrent cancers. With the increasing number of surviving cancer patients, understanding their recurrence patterns may allow more OR studies in the analysis of surveillance policies. In addition, modeling management processes while incorporating the evolution of the disease, and optimizing screening policies while incorporating service system constraints present interesting and challenging directions for future research in modeling screening services.

In this review, we had the opportunity of observing a new research trend that aims to capture the effects of human behavior on screening policies. Most of these studies focus on the patient compliance, while only a few consider the behavior of the medical doctors. Both types of behavior may have a significant impact on how the clinical guidelines and screening programs should be formed. We believe that incorporating human behavior, both at patient and doctor level, in OR models presents a significant challenge, which may have serious impacts on the design of screening programs.

Acknowledgements The authors gratefully acknowledge the feedback they received from Oguzhan Alagoz, Turgay Ayer, Brian Denton, Fatih Safa Erenay, Paul Harper, Evin Uzun Jacobson, Ozge Karanfil, and two anonymous reviewers, which helped improve the chapter significantly.

Appendix

See Tables 12.2, 12.3, 12.4, 12.5.

Table 12.2 The probability of observing outcome o given that state is s and action a is taken in year t ($f^t(o|s,a)$)

	Do nothing n				Colonoscopy c			
State	Observation				Observation			
s	$-$	R	$+$	$++$	$-$	R	$+$	$++$
H	1	0	0	0	1	0	0	0
P	1	0	0	0	$1\text{-}\alpha_P$	α_P	0	0
C	$1\text{-}\omega$	0	0	ω	$(1\text{-}\alpha_C)(1\text{-}\omega)$	0	α_C	$(1\text{-}\alpha_C)\omega$

Table 12.3 The transition probability to state s' when outcome o is observed given that state is s and action a is taken in year t ($p^t(s'|s,a,o)$)

	Observation								
	−				++ or +		R		
s/s'	H	P	C	D	T	D	H	P	D
H	$\lambda_H^t(a)$ $(1-\rho_{HP}^t)$	$\lambda_H^t(a)\rho_{HP}^t$	0	$1-\lambda_H^t(a)$	–	–	–	–	–
P	0	$\lambda_P^t(a)$ $(1-\rho_{PC}^t)$	$\lambda_P^t(a)\rho_{PC}^t$	$1-\lambda_P^t(a)$	–	–	$\lambda_P^t(c)$ $(1-\rho_{HP}^t)$	$\lambda_P^t(c)\rho_{HP}^t$	$1-$ $\lambda_P^t(c)$
C	0	0	$\lambda_C^t(a)$	$1-\lambda_C^t(a)$	$\lambda_C^t(a)$	$1-\lambda_C^t(a)$	–	–	–

Table 12.4 Transition probability from a state s to state s' when action n is taken in year t ($g^t(s'|s,n)$)

	State s'				
State s	H	P	C	T	D
H	$\lambda_H^t(n)(1-\rho_{HP}^t)$	$\lambda_H^t(n)\rho_{HP}^t$	0	0	$1-\lambda_H^t(n)$
P	0	$\lambda_P^t(n)(1-\rho_{PC}^t)$	$\lambda_P^t(n)\rho_{PC}^t$	0	$1-\lambda_P^t(n)$
C	0	0	$\lambda_C^t(n)(1-\omega)$	$\lambda_C^t(n)\omega$	$1-\lambda_C^t(n)$

Table 12.5 Transition probability from a state s to state s' when action c is taken in year t ($g^t(s'|s,c)$)

	State s'				
State s	H	P	C	T	D
H	$\lambda_H^t(c)(1-\rho_{HP}^t)$	$\lambda_H^t(c)\rho_{HP}^t$	0	0	$1-\lambda_H^t(c)$
P	$\lambda_P^t(c)(1-\rho_{HP}^t)\alpha_P$	$\lambda_P^t(c)\rho_{HP}^t\alpha_P+$ $\lambda_P^t(c)(1-\rho_{PC}^t)$ $(1-\alpha_P)$	$\lambda_P^t(c)\rho_{PC}^t(1-\alpha_P)$	0	$1-\lambda_P^t(c)$
C	0	0	$\lambda_C^t(c)(1-\omega)$ $(1-\alpha_C)$	$\lambda_C^t(c)\alpha_C+$ $\lambda_C^t(c)\omega(1-\alpha_C)$	$1-\lambda_C^t(c)$

References

Akhundov N (2015) Optimal location, patient routing, and capacity decisions for endoscopy clinical network in western ontario: a simulation-based optimization approach. MS Thesis, University of Waterloo

Alagöz O (2014) Chapter 6: Optimizing cancer screening using partially observable Markov decision processes. INFORMS, Catonsville, pp 75–89

Alagöz O, Ayer T, Erenay FS (2011) Operations research models for cancer screening. In: Cochran JJ, Cox LA Jr, Keskinocak P, Kharoufeh JP, Smith JC (eds) Wiley encyclopedia of operations research and management science. Wiley, Hoboken

Alistar S.S., Long EF, Brandeau ML, Beck EJ (2013) Hiv epidemic control—a model for optimal allocation of prevention and treatment resources. Health Care Manag Sci 17(2):162–181

American Cancer Society (2016) Colorectal cancer early detection: acs recommendations. http://www.cancer.org/cancer/colonandrectumcancer/

Andermann A, Blancquaert I, Beauchamp S (2008) Revisiting wilson and jungner in the genomic age: a review of screening criteria over the past 40 years. Bull World Health Organ 86:317–319

Ayer T (2015) Inverse optimization for assessing emerging technologies in breast cancer screening. Ann Oper Res 230:57–85

Ayer T, Alagöz O, Stout NK (2012) Or forum: a pomdp approach to personalize mammography screening decisions. Oper Res 60(5):1019–1034

Ayer T, Alagöz O, Stout NK, Burnside ES (2016) Heterogeneity in womens adherence and its role in optimal breast cancer screening policies. Management Science 62(5):1339–1362

Ayvacı MUS, Alagöz O, Burnside ES (2012) The effect of budgetary restrictions on breast cancer diagnostic decisions. Manuf Serv Oper Manag 14(4):600–617

Baker RD (1998) Use of a mathematical model to evaluate breast cancer screening policy. Health Care Manag Sci 1(2):103–113

Brailsford SC, Davies R, Canning C, Roderick PJ (1998) Evaluating screening policies for the early detection of retinopathy in patients with noninsulin dependent diabetes. Health Care Manag Sci 1(2):115–124

Brailsford SC, Gutjahr WJ, Rauner MS, Zeppelzauer W (2006) Combined discrete-event simulation and ant colony optimisation approach for selecting optimal screening policies for diabetic retinopathy. Comput Manag Sci 4(1):59–83

Brailsford S, Harper P, Sykes J (2012) Incorporating human behaviour in simulation models of screening for breast cancer. Eur J Oper Res 219:491–507

Brandeau ML (2013) OR in public health: a little help can go a long way. In: Operations research and health care policy. Springer International Publishing, New York

Brandeau ML, Zaric GS (2009) Optimal investment in hiv prevention programs: more is not always better. Health Care Manag Sci 12(2):27–37

Brandeau ML, Owens DK, Sox CH, Wachter RM (1993) Screening women of childbearing age for human immunodeficiency virus: a model-based policy analysis. Manag Sci 39(1):72–92

Chen Q, Ayer T, Chhatwal J (2016) Optimal liver cancer surveillance in hepatitis C-infected population (working paper)

Chhatwal J, Alagoz O, Burnside ES (2010) Optimal breast biopsy decision-making based on mammographic features and demographic factors. Oper Res 58(6):1577–1591

CISNET (2016) Cisnet model registry. https://resources.cisnet.cancer.gov/registry/packages/

Davari S, Kılıç K, Naderi S (2016) A heuristic approach to solve the preventive health care problem with budget and congestion constraints. Appl Math Comput 276:442–453

de Kok IMCM, van Rosmalen J, Dillner J, Arbyn M, Sasieni P, Iftner T, van Ballegooijen M (2012) Primary screening for human papillomavirus compared with cytology screening for cervical cancer in european settings: cost effectiveness analysis based on a dutch microsimulation model. BMJ 344:e670

Eagle JN (1984) The optimal search for a moving target when the search path is constrained. Oper Res 32(5):1107–1115

Eddy DM (1980) Screening for cancer: theory analysis and design. Prentice-Hall, Inc., Englewood Cliffs

Eddy DM (1983) Optimal incentives for allocating hiv/aids prevention resources among multiple populations. Med Decis Mak 3(1):45–62

Eddy DM (1990) Screening for colorectal cancer. Ann Intern Med 113(5):373–384

Erenay FS, Alagoz O, Said A (2014) Optimizing colonoscopy screening for colorectal cancer prevention and surveillance. Manuf Serv Oper Manag 16(3):381–400

Erenay FS, Alagoz O, Banerjee R, Said A, Cima RR (2016) Cost-effectiveness of alternative colonoscopy surveillance strategies to mitigate metachronous colorectal cancer incidence. Cancer 122(16):2560–2570

Food and Drug Administration (2011) Compliance guidance: the mammography quality standards act final regulations: preparing for mqsa inspections; final. http://www.fda.gov/RegulatoryInformation/Guidances/ucm094420.htm

Frazier A, Colditz G, Fuchs C, Kuntz K (2000) Cost-effectiveness of screening for colorectal cancer in the general population. JAMA 284(15):1954–1961

Grosse SD (2008) Assessing cost-effectiveness in healthcare: history of the $50,000 per qaly threshold. Expert Rev Pharmacoecon Outcomes Res 8(2):165–178

Gu W, Wang X, McGregor SE (2010) Optimization of preventive health care facility locations. Int J Health Geogr 9(1):1–16

Gulati R, Gore JL, Etzioni R (2013) Comparative effectiveness of alternative prostate-specific antigen-based prostate cancer screening strategies model: estimates of potential benefits and harms. Ann Intern Med 158(3):145–153

Güneş ED, Nickel S (2015) Location problems in healthcare. In: Location science. Springer International Publishing, Cham, pp 555–579

Güneş ED, Chick SE, Akşin OZ (2004) Breast cancer screening services: trade-offs in quality, capacity, outreach, and centralization. Health Care Manag Sci 7(4):291–303

Güneş ED, Örmeci EL, Kunduzcu D (2015) Preventing and diagnosing colorectal cancer with a limited colonoscopy resource. Prod Oper Manag 24(1):1–20

Habbema J, van Oortmarssen G, Lubbe J, van der Maas P (1985) The miscan simulation program for the evaluation of screening for disease. Comput Methods Prog Biomed 20(1):79–93

Harper P, Jones S (2005) Mathematical models for the early detection and treatment of colorectal cancer. Health Care Manag Sci 8:101–109

Heidenberger K (1996) Strategic investment in preventive health care: quantitative modeling for programme selection and resource allocation. OR Spektr 18:1–14

Ivy JS (2009) Can we do better? Optimization models for breast cancer screening. In: Handbook of optimization in medicine. Springer US, Boston, pp 1–28

Jones AP, Homer JB, Murphy DL, Essien JDK, Milstein B, Seville DA (2006) Understanding diabetes population dynamics through simulation modeling and experimentation. Am J Public Health 96(3):488–494

Karanfil, Ö (2016) Why clinical practice guidelines shift over time: a dynamic model with application to prostate cancer screening. PhD Thesis, Massachusetts Institute of Technology

Kirkizlar E, Faissol DM, Griffin PM, Swann JL (2010) Timing of testing and treatment for asymptomatic diseases. Math. Biosci. 26(1):28–37

Lee HL, Pierskalla WP (1988), Mass screening models for contagious diseases with no latent period. Oper Res 36(6):917–928

Lee E, Lavieri MS, Volk ML, Xu Y (2015) Applying reinforcement learning techniques to detect hepatocellular carcinoma under limited screening capacity. Health Care Manag Sci 18:363–375

Leshno M, Halpern Z, Arber N (2003) Cost effectiveness of colorectal cancer screening in the average risk population. Health Care Manag Sci 6(3):165–174

Li Y, Zhu M, Klein R, Kong N (2014) Using a partially observable Markov chain model to assess colonoscopy screening strategies–a cohort study. Eur J Oper Res 238(1):313–326

Li J, Dong M, Ren Y, Yin K (2015) How patient compliance impacts the recommendations for colorectal cancer screening. J Comb Optim 30:920–937

Lincoln T, Weiss GH (1964) A statistical evaluation of recurrent medical examinations. Oper Res 12(2):187–205

Loeve F, Boer R, van Oortmarssen GJ, van Ballegooijen M, Habbema JDF (1999) The miscan-colon simulation model for the evaluation of colorectal cancer screening. Comput Biomed Res 32:13–33

Lovejoy WS (1991) A survey of algorithmic methods for partially observed markov decision processes. Ann Oper Res 28(1):47–65

Macafee D, Waller M, Whynes D, Moss S, Scholefield J (2008) Population screening for colorectal cancer: the implications of an ageing population. Br J Cancer 99:1991–2000

Madadi M, Zhang S, Henderson LM (2015) Evaluation of breast cancer mammography screening policies considering adherence behavior. Eur J Oper Res 247:630–640

Maillart LM, Ivy JS, Ransom S, Diehl K (2008) Assessing dynamic breast cancer screening policies. Oper Res 56(6):1411–1427

Malvankar-Mehta MS, Xie B (2012) Optimal incentives for allocating hiv/aids prevention resources among multiple populations. Health Care Manag Sci 15(4):327–338

McLay LA, Foufoulides C, Merrick JRW (2010) Using simulation-optimization to construct screening strategies for cervical cancer. Health Care Manag Sci 13(4):294–318

McMahon PM, Kong CY, Bouzan C, Weinstein MC, Cipriano LE, Tramontano AC, Johnson BE, Weeks JC, Gazelle GS (2011) Cost-effectiveness of computed tomography screening for lung cancer in the united states. J Thorac Oncol 6(11):1841 – 1848

Monahan GE (1982) A survey of partially observable markov decision processes: theory, models, and algorithms. Manag Sci 28(1):1–16

Önen Z, Sayın S, Gurvit IH (2016) Optimal population screening policies for alzheimer's disease. Working Paper, Koç University

Örmeci EL, Güneş ED, Kunduzcu D (2015) A modeling framework for control of preventive services. Manuf Serv Oper Manag 18(2):227–244

Özekici S, Pliska SR (1991) Optimal scheduling of inspections: a delayed markov model with false positives and negatives. Oper Res 39(2):261–273

Parmigiani G (1996) Optimal scheduling of fallible inspections. Oper Res 44(2):360–367

Price, S, Golden B, Wasil E, Denton BT (2016) Operations research models and methods in the screening, detection, and treatment of prostate cancer: a categorized, annotated review. Manuf Serv Oper Manag 8:9–21

Rauner M, Gutjahr W, Heidenberger K, Wagner J, Pasia J (2010) Dynamic policy modeling for chronic diseases: metaheuristic-based identification of pareto-optimal screening strategies. Oper Res 58(5):1269–1286

Rutter CM, Zaslavsky A, Feuer E (2011) Dynamic microsimulation models for health outcomes: a review. Med Decis Mak 31(1):10–18

Shwartz M (1978) A mathematical model used to analyze breast cancer screening strategies. Oper Res 26(6):937–955

Siebert U, Sroczynski G, Hillemanns P, Engel J, Stabenow R, Stegmaier C, Voigt K, Gibis B, Hölzel D, Goldie SJ (2006) The german cervical cancer screening model: development and validation of a decision-analytic model for cervical cancer screening in germany. Eur J Pub Health 16(2):185–192

Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, Kuntz KM (2012) State-transition modeling: a report of the ispor-smdm modeling good research practices task force-3. Value Health 15(6):812–820

Smallwood RD, Sondik EJ (1973) The optimal control of partially observable markov processes over a finite horizon. Oper Res 21(5):1071–1088

Steimle LN, Denton BT (2017) Markov decision processes for screening and treatment of chronic diseases. In Markov Decision Processes in Practice (pp. 189–222). Springer International Publishing.

Szeto KL, Devlin NJ (1996) The cost-effectiveness of mammography screening: evidence from a microsimulation model for New Zealand. Health Policy 38(2):101–115

Tejada JJ, Ivy JS, King RE, Wilson JR, Ballan MJ, Kay MG, Diehl KM, Yankaskas BC (2014) Combined des/sd model of breast cancer screening for older women, ii: screening-and-treatment simulation. IIE Trans 46(7):707–727

van Ballegooijen M, Habbema JD, van Oortmarssen GJ, Koopmanschap MA, Lubbe JT, van Agt HM (1992) Preventive pap-smears: balancing costs, risks and benefits. Br J Cancer 65(6): 930–933

Verter V, Lapierre S (2002) Location of preventive health care facilities. Ann Oper Res 110(1):123–132

Verter V, Zhang Y (2015) Location models for preventive care. In: Applications of location analysis. Springer International Publishing, Cham, pp 223–241

Vieira IT, de Senna V, Harper PR, Shahani AK (2011) Tumour doubling times and the length bias in breast cancer screening programmes. Health Care Manag Sci 14(2):203–211

Voelker JA, Pierskalla WP (1980) Test selection for a mass screeening program. Nav Res Logist 27:43–55

Waugh N, Scotland G, McNamee P, Gillett M, Brennan A (2007) Screening for type 2 diabetes: literature review and economic modelling. Health Technol Assess 11(17):144

Wilson JMG, Jungner G (1968) Principles and practice of screening for disease. World Health Organization, Geneva

Wilson R, Liston J (2011) Quality assurance guidelines for breast cancer screening radiology, Volume NHSBSP Publication No 59. NHS Cancer Screening Programmes

Yang Y, Goldhaber-Fiebert JD, Wein LM (2013) Analyzing screening policies for childhood obesity. Manag Sci 59(4):782–795

Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM (2008) Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. preventive services task force. Ann Intern Med 149(9):659–669

Zelen M, Feinleib M (1969) On the theory of screening for chronic diseases. Biometrika 56(3):601–614

Zhang Y, Berman O, Verter V (2009) Incorporating congestion in preventive healthcare facility network design. Eur J Oper Res 198(3):922–935

Zhang Y, Berman O, Marcotte P, Verter V (2010) A bilevel model for preventive healthcare facility network design with congestion. IIE Trans 42(12):865–880

Zhang Y, Berman O, Verter V (2012) The impact of client choice on preventive healthcare facility network design. OR Spectr 34(2):349–370

Zhang J, Denton BT, Balasubramanian H, Shah ND, Inman BA (2012a) Optimization of prostate biopsy referral decisions. Manuf Serv Oper Manag 14(4):529–547

Zhang J, Denton BT, Balasubramanian H, Shah ND, Inman BA (2012b) Optimization of psa screening policies: a comparison of the patient and societal perspectives. Med Decis Mak 32(2):337–349

Zhang Y, Denton BT, Nielsen ME (2013) Comparison of surveillance strategies for low-risk bladder cancer patients. Med Dec Mak 33(2):198–214

Zhang J, Mason JE, Denton BT, Pierskalla WP (2013) Disease prevention, detection, and treatment. In: Cochran JJ, Cox LA Jr, Keskinocak P, Kharoufeh JP, Smith JC (eds) Encyclopedia of operations research and management science. Springer US, New York, pp 437–447.

Zimmerman S (1997) Factors influencing hispanic participation in prostate cancer screening. Oncol Nurs Forum 24:499–504

Chapter 13
Classification of Cancer Data: Analyzing Gene Expression Data Using a Fuzzy Decision Tree Algorithm

Simone A. Ludwig, Stjepan Picek, and Domagoj Jakobovic

13.1 Introduction

Data mining is the process of extracting useful information from the knowledge that is hidden in large volumes of data. The aim in data mining is to find patterns and relationships of data using data analysis tools and other techniques to build models. There are two distinct models in data mining: predictive models and descriptive models. The predictive models use data with known outcomes to develop a model that is then used to explicitly predict the different outcomes. The other model is the descriptive model, which is used to describe patterns in existing data. Both types of models provide an abstract representation of the data, which can then guide in the understanding of the data analyzed.

Data mining techniques have proved to be indispensable when working with large sets of data. The data mining community has been active in research of various techniques as well as new applications of data mining for more than 50 years. Naturally, during that time a plethora of techniques was designed to deal with various scenarios where one well known methodology is based on decision trees. We can trace the roots of its popularity to the fact that such methods can easily

S.A. Ludwig (✉)
Department of Computer Science, North Dakota State University, Fargo, ND, USA
e-mail: simone.ludwig@ndsu.edu

S. Picek
KU Leuven, ESAT/COSIC and iMinds, Kasteelpark Arenberg 10, bus 2452, B-3001 Leuven-Heverlee, Belgium
e-mail: stjepan@computer.org

D. Jakobovic
Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia
e-mail: domagoj.jakobovic@fer.hr

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_13

be interpreted by humans and the extracted knowledge can be clearly presented and visualized (Breiman et al. 1984). However, often we encounter problems where decision trees need to have a strict division between feature values in data sets. In order to deal with that, Fuzzy Decision Tree (FDT) algorithms emerged (Chang and Pavlidis 1977). This chapter investigates the improvements in classification accuracy that fuzzy decision trees may exhibit compared to classical decision tree algorithms.

When discussing the areas where data mining techniques play an important role, the biomedical domain is doubtless a prominent one. Here, the data can be various measurements taken from patients (e.g. heart rhythm or electrocardiogram) or the genes themselves. In order to query the expression of a multitude of genes, gene expression profiling is used. It presents the measurement of the activity of a large number of genes at once in order to be able to verify the cellular function. When the focus is on cancer data sets, gene expression profiling is used to more accurately classify tumors. Besides classifying tumors, with more powerful gene expression techniques it is also possible to classify tumor subclasses.

The objective of these methods is to discover not only a single association but several associations of genes. For this purpose, many features must be considered, with typically very few of them being significant for any given classification. Additionally, relatively few data points are available for learning.

Although very popular in practice, classical decision trees share some disadvantages that are revealed under these conditions. Specifically, their performance tends to deteriorate with the increase of features and emergence of complex interactions. Since most decision trees divide the search space into mutually exclusive regions, often the resulting tree must include several copies of the same subtree to accurately represent the data. Furthermore, their greedy behavior is prone to over-fitting to the training set, as well as irrelevant features and noise.

In contrast to that, fuzzy decision trees do not need to assign a data instance with a single branch, but may simultaneously assign more branches to the same instance with a gradual certainty. In this way, fuzzy decision trees retain the symbolic tree structure, but are able to represent concepts by producing continuous classification outputs with gradual transitions between classes.

In this work, we experiment with a fuzzy decision tree algorithm with the goal of analyzing gene expression cancer data. Besides the comparison with a decision tree algorithm, we also compare the proposed algorithm with several other well known algorithms for classification. The results present the advantages of fuzzy decision trees over classical decision trees for multiple data sets in this domain.

This chapter is an extended version of the paper published in Ludwig et al. (2015), and is arranged as follows: Section 13.2 describes the related work. The proposed approach is introduced in Sect. 13.3. The experimental setup and results are demonstrated in Sect. 13.4. In the final section (Sect. 13.5) the conclusions of this research are discussed.

13.2 Related Work

We divide the relevant research into two categories; the first is concerned with fuzzy decision tree development and applications, and the second with the applications of data mining techniques in the analysis of medical data. However, since this still encompasses a huge research area, we concentrate only on a subset of papers exploring cancer data research.

The development of fuzzy variants of decision tree induction has been around for quite a while (Chang and Pavlidis 1977; Janikow 1998), but they become a topic of interest in recent applications. These approaches provide examples for the application of "fuzzification" to standard machine learning methods.

There are many variations of fuzzy decision trees. Soft Decision Trees (SDT) are presented in Olaru and Wehenkel (2003), which combine tree-growing and pruning to determine the structure and refitting and backfitting to improve the generalization capability. The authors empirically show that SDTs are more accurate than standard decision trees. In (An and Hu 2012), the authors propose fuzzy-rough classification trees with a new measure to quantify the functional dependency of decision attributes on condition attributes within fuzzy data. The experiments show that fuzzy-rough classification trees outperform existing decision tree induction algorithms on 16 real-world datasets.

Fuzzy decision trees have been applied to various domains; in Chang et al. (2010) they are integrated with genetic algorithms for data classification in database applications, and in Lai et al. (2009) for developing a financial time series-forecasting model, where they were also combined with a genetic algorithm.

In (Biswal and Dash 2013), the authors use a FDT-based classifier for the measurement, identification, and classification of various types of power quality disturbances and they report robust performance under different noise conditions. A fuzzy knowledge-based network is developed in Mitra et al. (2002) based on the linguistic rules extracted from a fuzzy decision tree. The effectiveness of the system, in terms of recognition scores, structure of decision tree, performance of rules, and network size, is extensively demonstrated on three sets of real-life data.

For the biomedical applications, we first enumerate several surveys on the data mining techniques and cancer data. In the scope of cancer data analysis, a survey with a comprehensive study of various cancer classification methods is given in Lu and Han (2003). The authors conduct an analysis of the efficiency of methods based on their speed, accuracy and ability to reveal biologically meaningful gene information. Another survey on data mining techniques and breast cancer data is given in Padmapriya and Velmurugan (2014). In their work, the authors discuss the algorithms ID3 and C4.5. In (Palivela et al. 2013), the authors compare several data mining techniques on breast cancer data. A survey on decision tree classifiers in gene micro array data analysis is given in Polaka et al. (2010). A general framework of sample weighting to improve the stability of feature selection methods is proposed in Yu et al. (2012).

Experimentation with a multiclass classifier based on SVM (Support Vector Machine) algorithm is reported in Ramaswamy et al. (2001). The authors use samples of 14 common tumor types and achieve an overall classification accuracy of 78%. A method of gene selection with reliability analysis is devised in order to help differentiate between histologically similar cancers (Li and Casey 2004). In (Cuperlovic-Culf et al. 2005), the question is addressed on how to correctly select diagnostic marker genes from the gene expression profiles.

New astrocytic tumor micro-array gene expression data set is experimented with using an artificial neural network algorithm (Petalidis et al. 2008). With this algorithm the authors address grading of human astrocytic tumors, derive specific transcriptional signatures from histopathologic subtypes of astrocytic tumors, and assess whether these molecular signatures define survival prognostic subclasses. Another artificial neural networks approach for classifying cancers to specific diagnostic categories based on their gene expression signatures is provided in (Khan et al. 2001).

DNA micro-array analysis with supervised classification has shown to identify a gene expression signature to be strongly predictive of a short interval to distant metastases for breast cancer patients (Veer et al. 2002). With this strategy it is possible to select the patients who would benefit from chemotherapy or hormonal therapy. The problem how to select a small subset of genes from large patterns of data recorded on DNA micro-arrays is addressed in Guyon et al. (2002). The authors experiment with SVM algorithms based on recursive feature elimination.

Another novel method called decision trunks that is based on decision trees to classify cancer using expression data is proposed in Ulfenborg et al. (2013). The results suggest that the new algorithm performs at least as good as the state of the art algorithms when considering accuracy.

The use fuzzy decision trees to predict breast cancer survivability is reported in Khan et al. (2008). The authors compare decision trees and fuzzy decision trees and find FDT to be more robust and balanced than DT. A logistic regression and decision trees for survivability prognosis in patients with breast cancer is given in Wang et al. (2013). The authors show that logical regression has better statistical power in predicting 5-year survivability.

In (Hamdan and Garibaldi 2010), an adaptive fuzzy inference system technique for the estimation of survival prediction in cancer patients is proposed. Three methods, namely, decision trees, artificial neural networks, and logistic regression to develop prediction models for breast cancer survivability is given in Delen et al. (2005). The authors found decision trees to be the predictor with the best accuracy.

13.3 Fuzzy Decision Tree Classifier

Supervised classification is a very important and frequently used technique that is applied in the area of medical informatics. The most commonly used classification algorithms include logic-based algorithms, neural network algorithms,

statistical learning algorithms, instance-based learning algorithms, and support vector machine algorithms.

In terms of learning-based models, there are two groups: decision trees and rule-based classifiers. Decision trees classify instances by sorting them based on feature values. A decision tree classifier builds a decision tree model that can be used for the classification of unseen data. The decision tree model consists of a series of observations (branch nodes) that lead to conclusions (leaf nodes). The main difference between classical decision tree modeling and fuzzy decision tree modeling is the use of crisp or soft discretization, respectively. Classical decision tree modeling uses crisp discretization, whereby the decision space is partitioned into a set of non-overlapping subspaces using the crisp discretization method. For soft discretization, the decision space is partitioned into a set of overlapping subspaces. For both classical and fuzzy decision trees, each path from the root node to a leaf node represents a classification rule.

The algorithm of the FDT classifier starts by sorting the continuous values of a feature. It then produces a possible candidate "cut-point", and "fuzzifies" the "cut-point" by using an entropy evaluation function. This checking of the best "cut-point" is done recursively and is applied to all attributes. Once all attributes have been soft discretized, the attribute with minimum value is selected to generate two child branches and nodes. This steps repeats until one of the stopping criteria is met. A detailed description of the algorithm can be found in Ludwig et al. (2015); Chen and Ludwig (2013).

In order to show the decision trees that are generated by a DT and FDT classifier, a diabetes data set (obtained from the UCI repository (Frank and Asuncion 2010)) has been analyzed. The diabetes data set consists of 8 features, 768 instances and 2 classes. The decision trees generated by a classical DT (J48) (WEKA's J48 algorithms was used (Witten et al. 2011)) and our FDT (Java implementation) are shown in Figs. 13.1 and 13.2, respectively. What we can see is that both decision trees are roughly of equal complexity, but different decision trees were generated in terms of the features used.

13.4 Experiments and Results

The FDT was implemented in Java as outlined in the previous section. The classical decision tree algorithm used for comparison is WEKA's J48 decision tree implementation (Witten et al. 2011). Other algorithms based on naive Bayes, Bayesian network, logistic regression, radial basis function neural network, and support vector machine are also used and compared with. All algorithms are further introduced in one of the following subsections.

In addition, since feature selection is a normal preprocessing step in data mining, WEKA's attribute selection method is used to filter out the relevant features. Results of both, FDT and J48, are given for the complete data set (all features) as well as the reduced feature set selected by the attribute selection method. Tenfold cross-validation was used for the training and testing of all experiments.

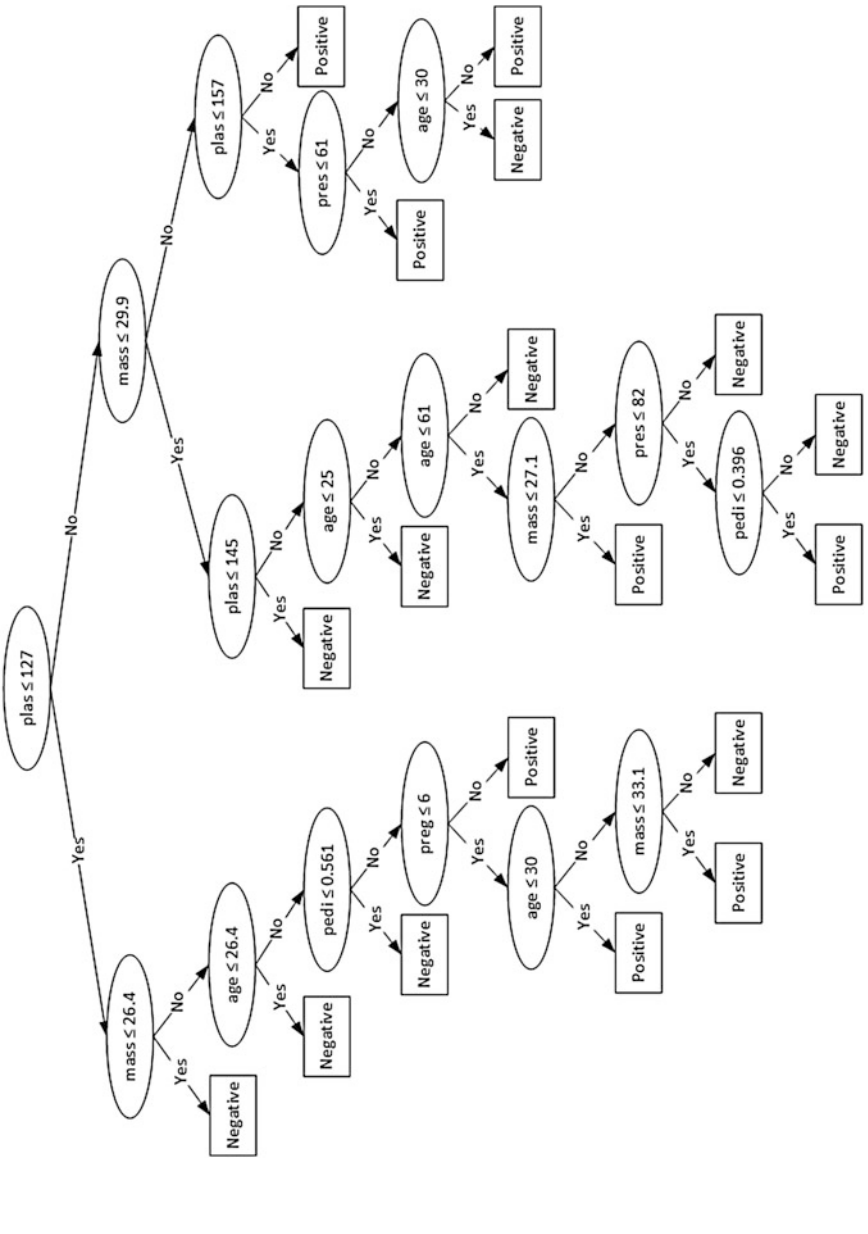

Fig. 13.1 Decision tree obtained from FDT classifier for the Ovarian cancer data set

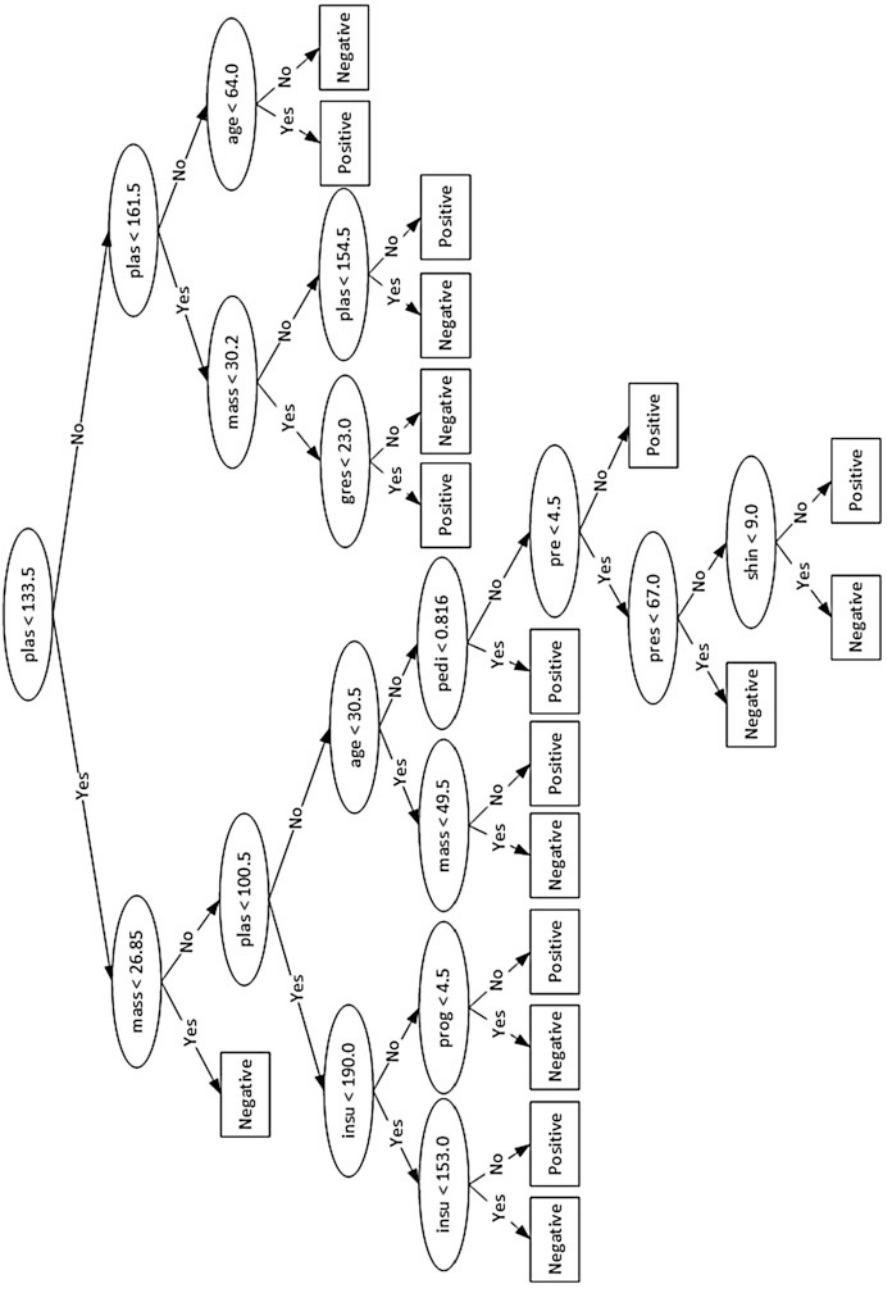

Fig. 13.2 Decision tree obtained from FDT classifier for the Prostate data set

13.4.1 Data Sets

The data sets[1] that have been chosen for this investigation are listed in Table 13.1. All data sets contain gene data information for different types of cancer. The number of features (all numeric) for the original data set (full) as well as after feature selection is applied is also given (reduced) in the column. The number of instances and the class balance of the binary data sets are also listed. Furthermore, a short description is provided and more details can be found looking up the references listed in the last column.

13.4.2 Evaluation Measures

In order to evaluate the medical data sets, the following measures have been chosen based on the number of True Positives (*TP*), True Negatives (*TN*), False Positives (*FP*), and False Negatives (*FN*):

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}. \tag{13.1}$$

$$\text{Sensitivity} = \frac{TP}{TP + FP}. \tag{13.2}$$

$$\text{Specificity} = \frac{TN}{FP + TN}. \tag{13.3}$$

Another measure used to evaluate medical data sets is the Receiver Operating Characteristic (ROC) (Swets 1996) curve, which is said to be a good indicator of the relationship between sensitivity and specificity. The AUC (Area Under the Curve) is calculated as follows:

$$\text{AUC} = \frac{1 - (1 - \text{Specificity}) + \text{Sensitivity}}{2}. \tag{13.4}$$

13.4.3 Comparison Algorithms

The implemented FDT algorithm is compared with a classical decision algorithm known as J48 (Quinlan 1993), which is implemented in WEKA. J48 is an extension of the C4.5 and the earlier ID3 algorithm (Quinlan 1979).

[1] http://datam.i2r.a-star.edu.sg/datasets/krbd/.

Table 13.1 Details of binary data sets used for experiments

Data set name	# of features	# of instances	Class balance	Size	Short description	Ref.
Colon tumor	full: 2000 reduced: 26	62	40/22	1.2 MB	Data collected from colon-cancer patients; tumor biopsies showing tumors ("negative"), and normal ("positive") biopsies are from healthy parts of colons of the same patients	(Alon et al. 1999)
Leukemia	full: 7129 reduced: 81	72	47/25	2.2 MB	Data collected from bone marrow samples; distinction is between Acute Myeloid Leukemia ("AML"), and Acute Lymphoblastic Leukemia ("ALL") without previous knowledge of these classes	(Golub et al. 1999)
Lung cancer	full: 12,533 reduced: 160	181	150/31	12 MB	Data collected from tissue samples; classification between Malignant Pleural Mesothelioma ("MPM"), and ADenoCArcinoma ("ADCA") of the lung	(Gordon et al. 2002)
Ovarian cancer	full: 15,154 reduced: 35	253	162/91	34 MB	Data to identify proteomic patterns in serum that distinguish ovarian cancer ("cancer") from non-cancer ("normal")	(Petricoin et al. 2002)
Prostate cancer	full: 12,600 reduced: 75	136	77/59	5.5 MB	Data from prostate tumor samples, whereby the non-tumor ("normal") prostate samples, and tumor samples ("cancer") are identified using 12,600 genes	(Singh et al. 2002)

The other comparison algorithms that are used for this investigation are:

- **NB**: is a Naive Bayes classifier implementation using estimator classes, whereby numeric estimator precision values are chosen based on the analysis of the training data.
- **BN:** implements a Bayes Network learning algorithm that uses various search algorithms and quality measures.
- **Log:** is a logistic regression model classifier. The classifier is based on a multinomial logistic regression model with a ridge estimator.
- **RBF:** is a radial basis function neural network model classifier. The classifier normalizes all attributes, and the initial centers for the Gaussian radial basis functions are identified using k-means.
- **SMO:** implements the sequential minimal optimization algorithm for training a support vector classifier. All missing values are replaced and nominal attributes are transformed into binary ones. In addition, all attributes are normalized by default.
- **BG:** implements the Bagging algorithm, which is an ensemble meta-algorithm that improves the accuracy and stability of learning algorithms that are used for classification and regression tasks.
- **RotF:** is the abbreviation for the Rotation forest algorithm that is a combination of decision trees with binary partitioning. Each decision tree is created based on the subset of training data with a bootstrap sample method.
- **RanF:** implements the Random forest algorithm. RanF uses a combination of decision trees with binary partitioning. Each tree is created based on training data with bootstrap sampling.

13.4.4 Experimental Results

Table 13.2 shows the accuracy, sensitivity and specificity values of the data sets using the complete feature set, i.e., using the complete data sets with all features. We can see that in terms of accuracy, the Ovarian cancer data sets achieves the highest values closely followed by the lung data set. However, comparing both data sets in terms of sensitivity and specificity reveals that the Ovarian cancer data set performs better scoring in the lower 90%.

Table 13.3 shows the same measures as Table 13.2, however, this time the feature set of the data sets are reduced after feature/attribute selection has been applied.

Table 13.2 Results of FDT measures with full feature set

Data set	Accuracy	Sensitivity	Specificity
Colon tumor	0.7746	0.8409	0.7200
Leukemia	0.8250	0.8475	0.6400
Lung cancer	0.9553	0.7879	0.9539
Ovarian cancer	0.9589	0.9175	0.9470
Prostate cancer	0.7985	0.8571	0.7885

Table 13.3 Results of FDT measures with reduced feature set

Data set	Accuracy	Sensitivity	Specificity
Colon tumor	0.8028	0.8864	0.7826
Leukemia	0.8750	0.8983	0.7391
Lung cancer	0.9553	0.7879	0.9540
Ovarian cancer	0.9711	0.9485	0.9662
Prostate cancer	0.8836	0.7662	0.7188

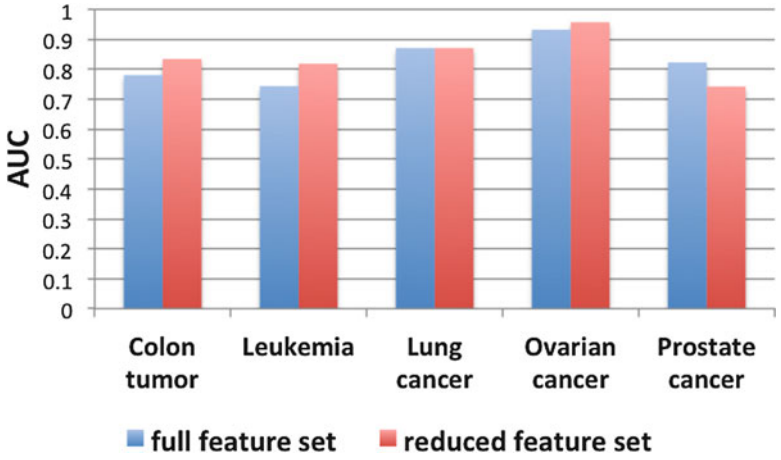

Fig. 13.3 Comparison of AUC values for different data sets with full and reduced feature set

We can see that the accuracy values are higher with the exception of the Lung cancer data set that scored the same accuracy. In terms of sensitivity and specificity, improved values can also be observed. Therefore, we can conclude that overall the feature reduction method improved the accuracy.

Figure 13.3 shows the AUC values for the data set with and without feature selection. The AUC values are often used since it shows the interplay between sensitivity and specificity. As can be seen by the figure, the AUC is higher for the reduced feature data sets with the exception of the Prostate cancer data set (Fig. 13.4).

Table 13.4 shows the accuracy values comparing FDT with J48 as well as showing the effect of using the complete data set with all the features versus using the reduced data set. As can be seen by the values in bold, on the full data set FDT outperformed J48 four out of five times, and on the reduced data sets FDT outperformed J48 three out of five times.

Table 13.5 shows the comparison of FDT, J48, the naive Bayes classifier (NB), the Bayesian network algorithm (BN), the logistic regression (Log), radial basis function network (RBF), and the support vector machine algorithm (SMO).

Based on the five data sets, the SMO algorithm performs best out of all classifiers. It scores best 7 out of 10 times when applied to the full data sets as well as the reduced data sets. SMO is closely followed by NB and BN (both scoring best 4

```
=== Classifier model (full training set) ===
SMO
Kernel used: Linear Kernel: K(x,y) = <x,y>
Classifier for classes: negative, positive
BinarySMO
Machine linear: showing attribute weights, not support vectors.
-0.2258 * (normalized) attribute143
+ 0.8376 * (normalized) attribute249
+ -0.237 * (normalized) attribute258
+ -0.4451 * (normalized) attribute279
+ 1.1883 * (normalized) attribute377
+ -0.1269 * (normalized) attribute467
+ -1.0661 * (normalized) attribute576
+ -0.5733 * (normalized) attribute625
+ -0.7617 * (normalized) attribute682
+ -0.5918 * (normalized) attribute763
+ 0.9659 * (normalized) attribute765
+ 0.2894 * (normalized) attribute897
+ -0.8163 * (normalized) attribute1042
+ -0.6559 * (normalized) attribute1153
+ -0.207 * (normalized) attribute1200
+ -0.1432 * (normalized) attribute1227
+ -0.5952 * (normalized) attribute1325
+ -0.0822 * (normalized) attribute1328
+ -0.529 * (normalized) attribute1412
+ 0.8739 * (normalized) attribute1423
+ 0.6139 * (normalized) attribute1560
+ -0.4861 * (normalized) attribute1562
+ 0.175 * (normalized) attribute1635
+ -0.1088 * (normalized) attribute1671
+ -0.8822 * (normalized) attribute1772
+ 0.2362 * (normalized) attribute1917
+ 0.0996
```

Fig. 13.4 WEKA's output of the model generated of the SMO classifier applied to the Lung cancer data set

Table 13.4 Results of comparison of FDT and J48 with full and reduced feature set

	Full feature set		Reduced feature set	
Data set	FDT	J48	FDT	J48
Colon tumor	0.7746	**0.8226**	0.8028	**0.8710**
Leukemia	**0.8250**	0.7917	**0.8750**	0.8472
Lung cancer	**0.9553**	0.9503	0.9553	**0.9613**
Ovarian cancer	**0.9594**	0.9565	**0.9711**	0.9605
Prostate cancer	**0.7985**	0.7941	**0.8836**	0.8824

times). In particular, SMO achieves 100% accuracy on the Lung cancer data set and the Ovarian cancer data set. The overall conclusions that can be drawn are that the SMO clearly outperforms all other classifiers including FDT and J48. FDT only achieves close results on the Lung and Ovarian data sets.

Table 13.5 Results of comparison of FDT with other WEKA algorithms in terms of accuracy

Data set		FDT	J48	NB	BN	Log	RBF	SMO	BG	RotF	RanF
Colon tumor	full	0.7746	0.8225	0.5323	0.7581	0.7097	0.7903	**0.8548**	0.7903	0.7742	0.7581
	reduced	0.8028	0.8710	0.8548	**0.9032**	0.7581	0.8710	0.8548	0.8710	0.8871	0.8226
Leukemia	full	0.8245	0.7917	**0.9861**	0.9722	0.9028	0.9306	**0.9861**	0.9028	0.9306	0.8750
	reduced	0.8750	0.8472	**1.0000**	**1.0000**	0.9583	**1.0000**	0.9861	0.8889	0.9583	0.9722
Lung cancer	full	0.9553	0.9503	0.9834	0.9834	0.9889	0.9779	**0.9945**	0.9779	0.9669	0.9834
	reduced	0.9553	0.9613	**1.0000**	**1.0000**	0.9945	0.9945	**1.0000**	0.9779	0.9890	**1.0000**
Ovarian cancer	full	0.9594	0.9565	0.9249	0.9210	0.9841	0.8340	**1.0000**	0.9723	0.9658	0.9605
	reduced	0.9711	0.9605	**1.0000**	0.9960	**1.0000**	**1.0000**	**1.0000**	0.9723	**1.0000**	0.9881
Prostate cancer	full	0.7985	0.7941	0.5588	0.6618	0.8456	0.6617	**0.9118**	0.8529	0.9044	0.7941
	reduced	0.8836	0.8824	0.6176	**0.9559**	0.7647	0.7647	0.8676	0.8676	0.9412	0.9412

Investigating the generated models of the FDT algorithm in the form of decision trees (as shown in Figs. 13.5, 13.6, 13.7, 13.8 and 13.9) as compared to the best-performing SVM (SMO) classifier reveals that only a fraction of the features are used for the model of FDT, whereas all features are used for the SMO model generation (the reduced feature set is used). This is true for all other non decision tree algorithms. Table 13.6 lists the number of features of the model created by SMO and other algorithms, and FDT, respectively. For example, for the Colon tumor data set only 6 as compared to 26 features are used for the model of FDT versus all others including SMO, and even a wider gap is observed for the Lung cancer data set on which FDT uses 3 features whereas SMO and others use 160 features. This demonstrates that the FDT models are much simpler in terms of complexity as well as comprehensibility. To show an example of the models created by FDT and SMO, the model of a decision tree generated by FDT is shown in Fig. 13.7, and the model generated by SMO on the lung cancer data set is as given in Fig. 13.4 (output from WEKA console):

What can be observed by the comparison of the model generated by SMO versus the decision tree model generated by FDT, is the simple and easy to visualize and understand model that is generated by the decision tree model. The mathematical formula involving all attributes as given as the SMO model is more difficult to describe and interpret. Besides SMO, the other machine learning algorithms used for comparison involve a mathematical model generation that is similar in outcome than the SMO model.

To further discuss and interpret the generated decision trees, let us look at the decision tree generated for the Lung cancer data set (see Fig. 13.7). The constructed decision tree is based on three decision node, namely *1394_at*, *34320_at*, and *37716_at*. Given this decision tree, a unseen example can then be routed down the tree to reach a decision node in order to present the output. For example, if a patient has the following values: *1394_at=420*, *34320_at=2100*, and *37716_at=1500*, then the output will be *Mesothelioma*. The decision tree model is very intuitive since the resulting model is easy to understand and assimilate by humans. That is the reason for its popularity in particular in the medical domain.

13.5 Conclusion

This chapter investigated a fuzzy decision tree implementation applied to the classification of gene expression data. Five high-dimensionality cancer data sets were analyzed and compared with a classical decision tree algorithm as well as other well-known data mining algorithms.

The results revealed that comparing FDT with J48, the FDT algorithm outperformed J48 in terms of accuracy on four out of the five data sets when applied to the classification using the full data sets, and 3 out of 5 times when applied to the reduced data sets after feature selection was applied. In general, higher values of accuracy, sensitivity, and specificity were achieved on the preprocessed data sets as has been shown in past literature.

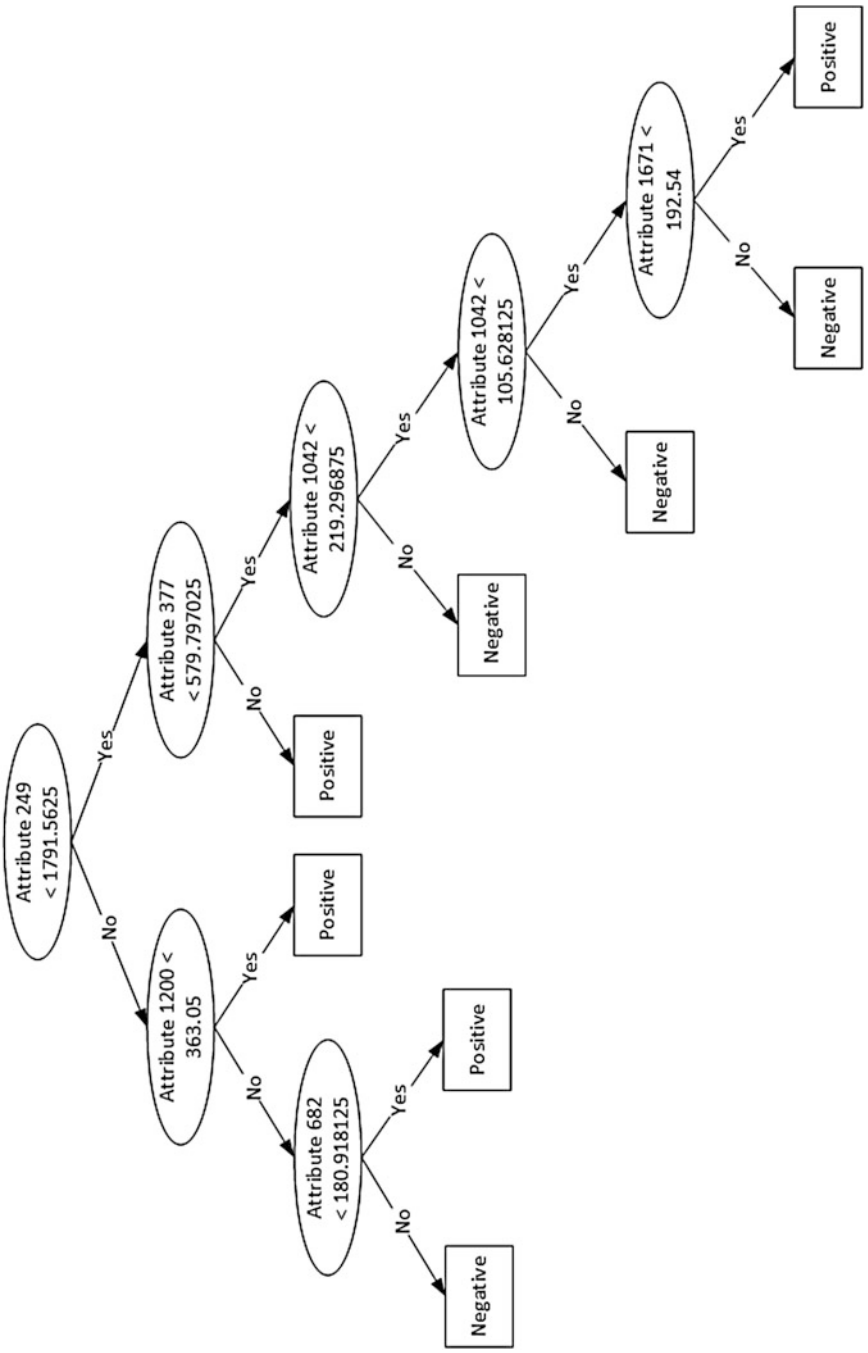

Fig. 13.5 Decision tree obtained from FDT classifier for the Colon tumor data set

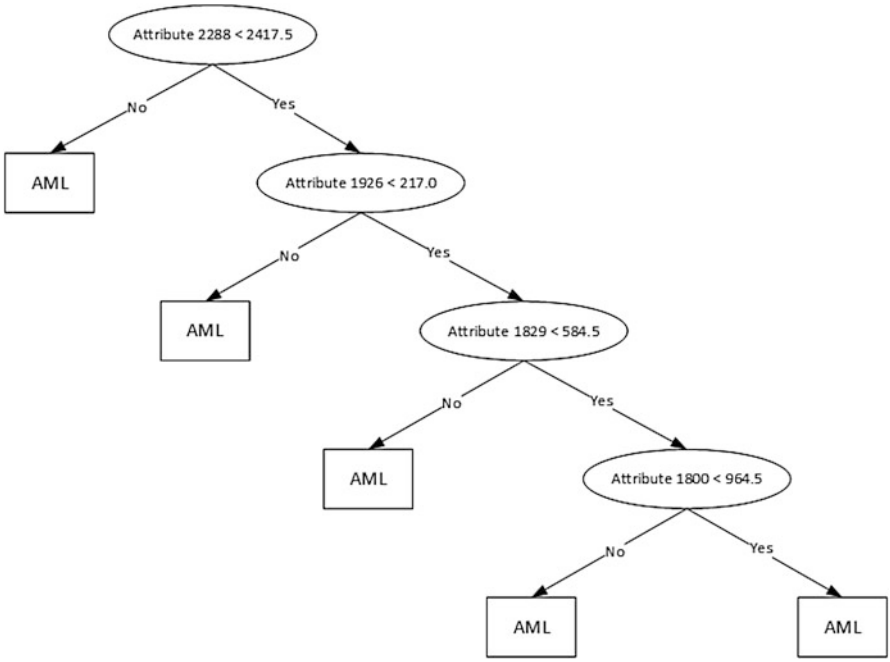

Fig. 13.6 Decision tree obtained from FDT classifier for the Leukemia data set

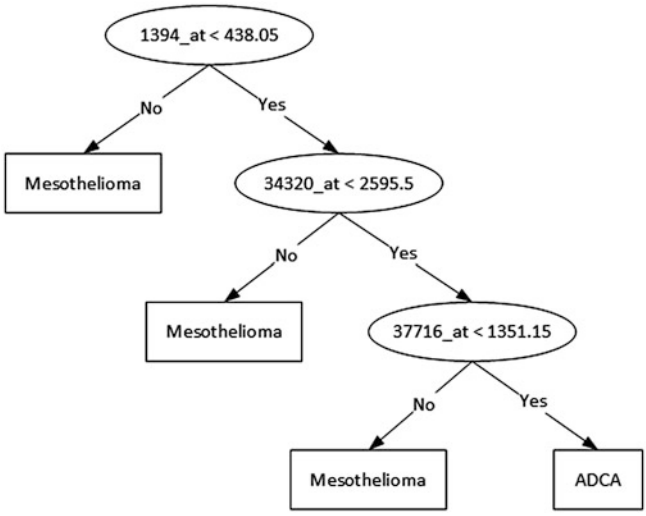

Fig. 13.7 Decision tree obtained from FDT classifier for the Lung cancer data set

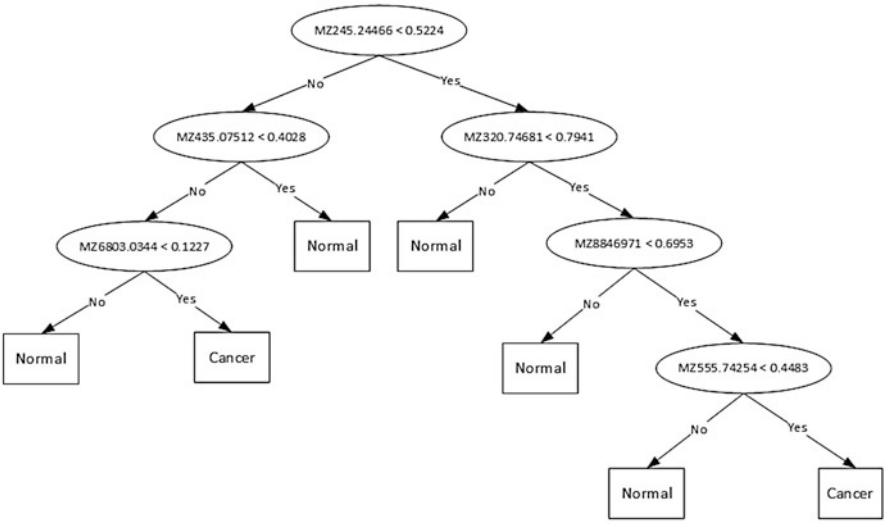

Fig. 13.8 Decision tree obtained from FDT classifier for the Ovarian cancer data set

Other measures of sensitivity and specificity were also in favor of FDT. The AUC values for FDT were also calculated and revealed that, in general, higher AUC values are achieved when the preprocessed data sets were investigated. In addition, the data sets, both full and reduced feature set, were run with common data mining algorithms and the support vector machine algorithm outperformed all other data mining algorithms achieving 100% accuracy on some data sets. This implies that the decision tree algorithms (both FDT and J48) are not the best choice when analyzing the five gene cancer data sets when accuracy is the only concern.

Further analyzing the complexity of the resulting models comparing the overall best-performing SVM algorithm with the FDT algorithm revealed that the model of FDT is many times less complex since only a fraction of features are used for FDT as compared to SVM, which uses all features. The compactness of the resulting decision tree model of FDT as well as the comprehensibility of the model are the strengths of the decision tree algorithms including the implemented FDT algorithm.

To summarize, the benefits of the decision tree model are: (1) in-build feature selection, (2) nonlinear relationships between parameters do not affect the tree performance, and (3) easy to interpret and explain.

Future work includes the evaluation of the FDT algorithm on larger gene expression data sets once they become available. Furthermore, a possible improvement of the FDT algorithm with, for example, another algorithm such as neural networks could be investigated.

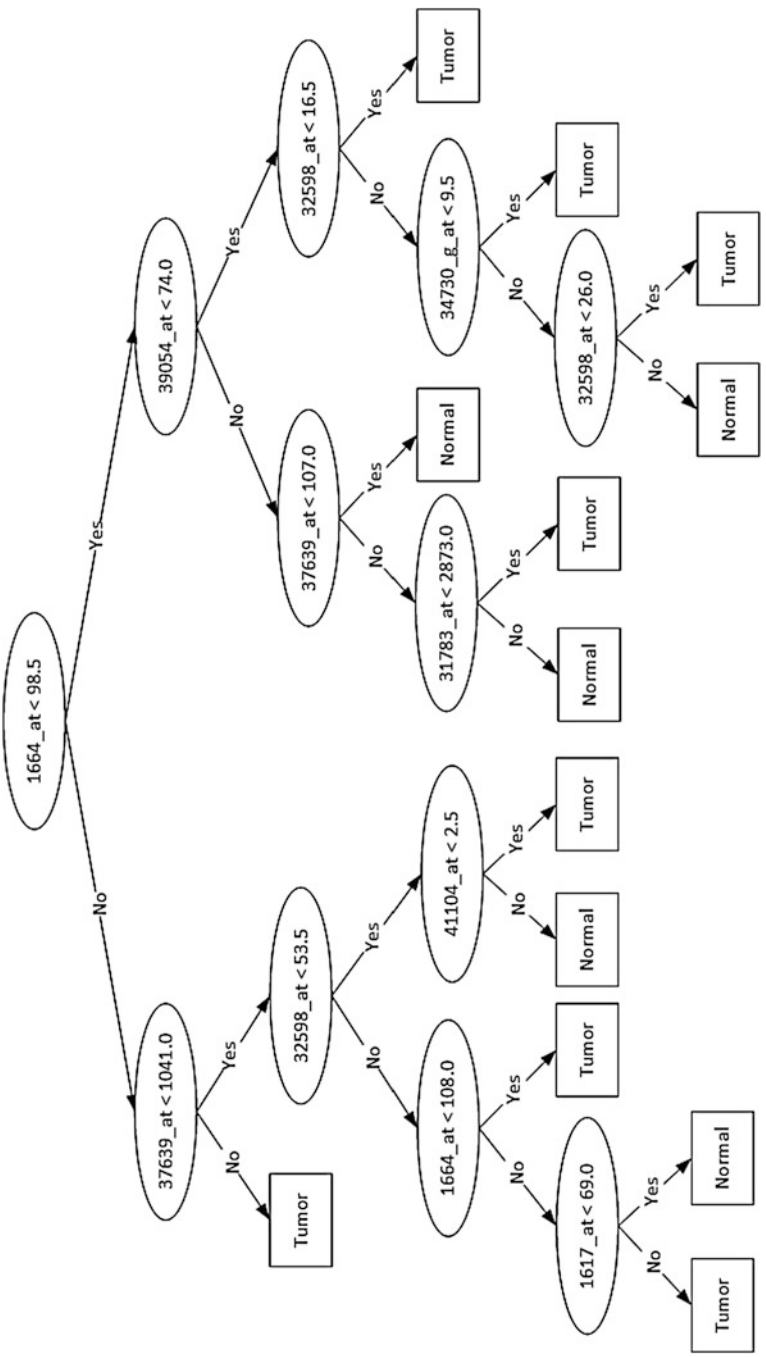

Fig. 13.9 Decision tree obtained from FDT classifier for the Prostate data set

Table 13.6 Comparison of features used for the generation of the model

	SMO and other algorithms	FDT
Colon tumor	26	6
Leukemia	81	4
Lung cancer	160	3
Ovarian cancer	35	6
Prostate cancer	75	8

Acknowledgements The authors would like to thank Gongyi Xia for the drawing of the figures of the decision trees.

References

Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci U S A 96(12):6745–6750

An S, Hu Q (2012) Fuzzy rough decision trees. In: Yao J et al (eds) Rough sets and current trends in computing, vol 7413. Springer, Berlin/Heidelberg, pp 397–404

Biswal M, Dash P (2013) Measurement and classification of simultaneous power signal patterns with an S-transform variant and fuzzy decision tree. IEEE Trans Ind Inf 9(4):1819–1827

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth, Belmont

Chang RL, Pavlidis T (1977) Fuzzy decision tree algorithms. IEEE Trans Syst Man Cybern 7(1):28–35

Chang P-C, Fan C-Y, Dzan W-Y (2010) A CBR-based fuzzy decision tree approach for database classification. Expert Syst Appl 37(1):214–225

Chen M, Ludwig SA (2013) Fuzzy decision tree using soft discretization and a genetic algorithm based feature selection method. In: World congress on nature and biologically inspired computing

Cuperlovic-Culf M, Belacel N, Ouellette, RJ (2005) Determination of tumour marker genes from gene expression data. Drug Discov Today10(6):429–437

Delen D, Walker G, Kadam A (2005) Predicting breast cancer survivability: a comparison of three data mining methods. Artif Intell Med 34(2):113–127

Frank A, Asuncion A (2010) UCI machine learning repository, University of california, school of information and computer science, Irvine, CA. http://archive.ics.uci.edu/ml

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286(5439):531–537

Gordon G, Jensen R, Hsiao L-L, Gullans S, Blumenstock J, Ramaswamy S, Richards WG, Sugarbaker DJ, Bueno R (2002) Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. Cancer Res 62:4963–4967

Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. Mach Learn 46(1–3):389–422

Hamdan H, Garibaldi J (2010) Adaptive neuro-fuzzy inference system (ANFIS) in modelling breast cancer survival. In: Fuzzy systems (fuzz), 2010 ieee international conference on, p 1–8

Janikow C (1998) Fuzzy decision trees: issues and methods. IEEE Trans Syst Man Cybern B Cybern 28(1):1–14

Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med 7(6):673–679

Khan MU, Choi JP, Shin H, Kim M (2008) Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare. In: Engineering in medicine and biology society, 2008. embs 2008. 30th annual international conference of the IEEE, pp 5148–5151

Lai RK, Fan C-Y, Huang W-H, Chang, P-C (2009) Evolving and clustering fuzzy decision tree for financial time series data forecasting. Expert Syst Appl 36(2, Part 2):3761–3773

Li MF, Casey SF (2004) Multi-class cancer subtype classification based on gene expression signatures with reliability analysis. FEBS Lett 561(13):186–190

Lu Y, Han J (2003) Cancer classification using gene expression data. Inf Syst 28:243–268

Ludwig SA, Jakobovic D, Picek S (2015) Analyzing gene expression data: fuzzy decision tree algorithm applied to the classification of cancer data. In: 2015 IEEE international conference on fuzzy systems

Mitra S, Konwar K, Pal S (2002) Fuzzy decision tree, linguistic rules and fuzzy knowledge-based network: generation and evaluation. IEEE Trans Syst Man Cybern Part C Appl Rev 32(4): 328–339

Olaru C, Wehenkel L (2003) A complete fuzzy decision tree technique. Fuzzy Sets Syst 138(2):221–254

Padmapriya B, Velmurugan T (2014) A survey on breast cancer analysis using data mining techniques. In: Computational intelligence and computing research (ICCIC), 2014 IEEE international conference on, pp 1–4

Palivela H, Yogish H, Vijaykumar S, Patil K (2013) Survey on mining techniques for breast cancer related data. In: Information communication and embedded systems (ICICES), 2013 international conference on, pp 540–546

Petalidis L, Oulas A, Backlund M, Wayland M, Liu L, Plant K, Happerfield L, Freeman TC, Poirazi P, Collins V (2008) Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data. Mol Cancer Ther 7(5):1013–1024

Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA (2002) Use of proteomic patterns in serum to identify ovarian cancer. The Lancet 359(9306):572–577

Polaka I, Tom I, Borisov A (2010) Decision Tree Classifiers in Bioinformatics. J. Riga Tech. Univ. 42:118–123

Quinlan JR (1979) Discovering rules by induction from large collections of examples. In: Michie D (ed) Expert systems in the microlectronic age. Edinburgh University Press, Edinburgh, pp 168–201

Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA

Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR (2001) Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci U S A 98(26):15149–15154

Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, Richie JP, Lander ES, Loda M, Kantoff PW, Golub TR, Sellers WR (2002) Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 1(2):203–209

Swets JA (1996) Signal detection theory and ROC analysis in psychology and diagnostics: collected papers. Lawrence Erlbaum Associates, Mahwah, NJ

Ulfenborg B, Klinga-Levan K, Olsson B (2013) Classification of tumor samples from expression data using decision trunks. Cancer Informat 12:53–66

van't Veer LJ, Dai H, van de Vijver MJ, He YD, Augustinus AMH, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415(6871):530–536

Wang K-J, Makond B, Wang K-M (2013) An improved survivability prognosis of breast cancer by using sampling and feature selection technique to solve imbalanced patient classification data. BMC Med Inform Decis Mak 13(1):124

Witten IH, Frank E, Hall MA (2011) Data mining: practical machine learning tools and techniques, 3rd ed. Morgan Kaufmann Publishers Inc., San Francisco, CA

Yu L, Han Y, Berens ME (2012) Stable gene selection from microarray data via sample weighting. IEEE/ACM Trans Comput Biol Bioinform 9(1):262–272

Part VI
Care Process: Treatment

Chapter 14
Efficiency of Diabetes Treatment

Evidence from the UK

Peter Wanke and Emel Aktas

14.1 Introduction

Diabetes is an emerging global epidemic linked to increases in physical inactivity, overweight, and obesity. According to the World Health Organization, 422 million adults have diabetes and 1.5 million deaths are directly attributed to this disease (WHO 2016). The total number of deaths from diabetes is expected to rise by more than 50% in the next decade with a notable increase by more than 80% in upper-middle income countries. DiabetesUK reports that the number of people with diabetes in the UK has reached an all-time high of 4 million, 0.5 million of whom are yet to be diagnosed (DiabetesUK 2016). This increases the pressure on the National Health Service (NHS) in particular when it comes to the inpatient care. Health care services in the UK are provided by the NHS, free of charge to patients. Although the original focus was the diagnosis and the treatment of the disease, now it has expanded to preventing ill health and improving the physical and mental health of the population (NHS 2013). The health care services are provided by a number of different organizations, such as clinical commissioning groups or charities. Primary care providers are responsible for physical and mental health and wellbeing in non-urgent cases.

The number of people living with diabetes has soared by nearly 60% in the past decade and 90% of cases are linked to diet and obesity, resulting in £869M spent on

P. Wanke
COPPEAD, The Federal University of Rio de Janeiro, Rua Pascoal Lemme, 355-Cidade Universitaria, 21941-918 Rio de Janeiro, Brazil
e-mail: peter@coppead.ufrj.br

E. Aktas (✉)
Cranfield School of Management, Cranfield University, College Road, MK43 0AL Cranfield, UK
e-mail: emel.aktas@cranfield.ac.uk

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_14

drugs, including insulin and metformin, an increase by 70% compared to 10 years ago (Gallagher 2015). On the other hand, Diabetes UK's annual review suggests very little overall improvement in diabetes care provision with some aspects of care getting worse, such as fewer people with type 1 diabetes receiving an annual check-up (Roberts 2015). That's why it is necessary to attempt at an efficiency evaluation of diabetes care providers and a predictive analysis to help suggest key actions for improving current care provision by benchmarking peer health care service providers.

In this chapter we analyze the efficiency of hospitals in England and Wales in terms of the diabetes inpatient care services using cross-sectional data published by the National Diabetes Inpatient Audit (NDIA) for the years 2012 and 2013. Each year an NDIA is delivered by the Health and Social Care Information Centre, covering hospitals and sites providing diabetes inpatient care in England and Wales, to establish whether diabetes management minimizes the risk of avoidable complications, the harm resulting from the inpatient stay, the patient experience of the inpatient stay, and the change in patient feedback on the quality of care since the beginning of the audit.

To assess the efficiency of hospitals/sites that provide diabetes inpatient care, we use the well-known TOPSIS method, due to the multi-criteria nature of the service. The data set comprises of 230 unique hospitals and sites that provide diabetes inpatient care with input measures about the process of care given to patients such as number of beds, nursing hours, consultant hours, dietitian hours, podiatrist hours, and pharmacist hours. Output measures comprise information about care outcomes and treatment such as the number of patients, percentage involved in treatment plan, percentage that is able to take control of the diabetes care, and overall satisfaction. We report what affects the efficiency of diabetes inpatient care and how the performance of hospitals and sites change from 2012 to 2013. We also build a neural network model to equip our efficiency assessment with predictive ability; so that we could provide meaningful suggestions to policy makers who would like to improve the diabetes care service in the best possible way.

Although, many studies assess several aspects of efficiency in health care, specific attention to diabetes treatment is limited. Our research is motivated by the lack of an efficiency analysis focusing on UK diabetes care facilities. The closest to our research focused on efficiency of primary diabetes care and analyzed 14 general practitioners (Amado and Dyson 2009). Our study is novel in its methodology—a two stage approach to efficiency analysis based on TOPSIS and neural networks—and its context, UK diabetes care facilities.

We propose a neural-network-based predictive model for efficiency of diabetes treatment in the UK based on the hospital demographics, patient demographics, diabetes type, medical routines, treatment dynamics and patient involvement, and the staff knowledge on diabetes, some of which are variables commonly used to assess healthcare service efficiency in the literature. For this purpose, we use the national diabetes inpatient audit, which is an annual snapshot of diabetes inpatient care in England and Wales (http://www.hscic.gov.uk/diabetesinpatientaudit). It is open to participation from hospitals with medical, surgical, gynecology wards or

intensive care units, allowing hospitals to benchmark diabetes care and to prioritize improvements in service provision.

Our methodology to achieve this objective comprises the use of TOPSIS and neural networks in a two-stage approach. We perform a TOPSIS analysis to assess the performance of diabetes health care service providers. We then use neural networks for building a predictive model of efficiency in the second stage as it predicts efficiency more flexibly and informatively than traditional statistical methods (Brockett et al. 1997). In Sect. 14.2 we present a state-of-the-art literature review. We explain the data and the two stage efficiency analysis and prediction model in Sect. 14.3. We discuss the results and policy implications in Sect. 14.4. We conclude the paper in Sect. 14.5.

14.2 Literature Review

Performance evaluation of health care facilities is relevant in all countries since most of the services have stretched budgets and it is gradually becoming more challenging to provide improved services without increasing costs. Efficiency analysis not only benchmarks the hospitals with respect to the best performers, but also identifies how the inefficient hospitals can improve their performance. Recently, Yang (2016) used a non-traditional DEA method to measure health indicators and allocate resources. Empirical results from Taiwan suggested that using this new method, resources can be allocated more efficiently than the existing method in the country; however, the author calls for models with a comprehensive set of efficiency indicators. Pelone et al. (2013) assessed the relative efficiency of primary care in Europe with a cross-sectional study performed on the data collected from 22 European countries in 2009 and 2010. Their key results informed policy makers that primary care quality depended on the coherence of systems rather than the quality of individual services provided where most efficient countries focused on access and coordination above everything else. Hollingsworth (2003) reviewed parametric and non-parametric methods to measure efficiency in health care services. Majority of the papers included in his review used DEA for efficiency analysis, followed by a combination of DEA and Tobit/regression models. Stochastic Frontier Analysis (SFA) was also reported to be an upcoming method to assess efficiency. The author called for more advanced methods in efficiency measurement in health sector with more emphasis on sensitivity analysis.

The efficiency analyses performed at a country level suggested that countries in which physicians were paid in wages and salaries and countries with capitation had higher efficiency than fee-for-service countries (Bhat 2005). In a similar line of research (Gok and Altındağ 2015) studied the impact of the pay for performance (PFP) system on the efficiencies of public and private hospitals and found that increased health care costs might reduce efficiency in private hospitals in contrast to public hospitals. The capitation payment model ensured additional funding for public hospitals as long as the quality of the service was at least similar to private hospitals (Gok and Altındağ 2015).

At healthcare service provider level, the efficiency could be assessed based on hospital resources and financial and activity outcomes, where the resources mainly involve hospital capacity, staff resources, and hospital expenses and the outcomes reflect emergency and regular visits (single-day care vs multiple-day care) as well as inpatient and outpatient healthcare service indicators (Hadji et al. 2014). In the systematic literature review on healthcare service efficiency, DEA and SFA were found to be the most frequently used efficiency analysis methods (Hadji et al. 2014). Nuti et al. (2016) compared university hospitals with general hospitals using 27 performance indicators in two stages comprising of a non-parametric Mann-Whitney U test in the first stage, followed by a robust equal variance test. The indicators chosen from existing performance evaluation systems are related to efficiency and appropriateness (e.g. relative stay index reflecting the length of stay), patient satisfaction (e.g. percentage of patients leaving the hospital against medical advice), economic and financial evaluation (e.g. average cost per weighted case), and outcomes (e.g. measures of 30-day mortality or readmissions for relevant inpatient activity). They cannot conclude whether university hospitals perform better than general hospitals due to heterogeneous results, but they call for a more integrated health care system with other facilities delivering community, primary, and outpatient care (Nuti et al. 2016).

Hospital efficiency is not a single-dimensional concept. Many factors contribute to the efficiency, which can be represented in terms of length of stay, number of patients, number of bed-days, etc. Multi Criteria Decision Making (MCDM) methods such as Analytical Hierarchy Process (AHP) (Saaty 1980; Creazza et al. 2016); Promethee (Brans and Vincke 1985; Corrente et al. 2013; Amaral and Costa 2014); Electre (Hatami-Marbini and Tavana 2011; Gomez and Carnero 2011); DEMATEL (Tsai 2009; Efe and Efe 2016), VIKOR (Opricovic and Tzeng 2007; Afful-Dadzie et al. 2016); TOPSIS (Lai et al. 1994; Wang et al. 2015) and UTA (Siskos et al. 2014; Grigoroudis and Zopounidis 2012) have found applications in efficiency assessment in health care. Among the MCDM methods, TOPSIS, which is developed by Hwang and Yoon (2012) in 1981, is a simple ranking method. TOPSIS uses attribute information, provides a cardinal ranking of alternatives, and does not require attribute preferences to be independent (Behzadian et al. 2012). In fact, the TOPSIS method can be used in a similar way to DEA to rank DMUs (Wang and Luo 2006; Wu 2006). While criteria weights are calculated as part of the DEA (Bouyssou 1999), in TOPSIS, they are determined with the decision maker (Wang et al. 2014). When many entities are on the efficient frontier, DEA will have low discriminatory power, whereas for TOPSIS this is not a problem (Wu 2006). Compared to SFA, neither DEA nor TOPSIS impose a functional form on the data or have any distributional assumptions for the efficiency scores (Wang et al. 2014).

Two-stage models combining efficiency assessment with efficiency prediction are emerging in the literature. In general, stage 1 is aimed assessing the efficiency of decision making units, whereas stage 2 is focused on assessing the impact of operational, environmental, or contextual variables on the efficiency scores. For example, (Mitropoulos et al. 2013) used DEA to assess the efficiency of each hospital in Greece, and then in the second stage of their analysis they assessed the

impact of the operational environment on the efficiency by regressing the efficiency scores on explanatory variables related to the performance of hospital services. Pelone et al. (2012) compared the efficiency of general practice delivered in various regions of Italy and tested whether the efficiency was affected by contextual factors using a two-stage methodology: assessing the efficiency using DEA and regressing the contextual factors to identify the most significant factors affecting the efficiencies. Tobit regression analysis showed a negative significant relationship between total public expenditure on health and efficiency, which was explained by the marginal utility of the money spent. The current literature on health care facilities could benefit from a more systematic and integrated approach to inform policy makes with the outputs of predictive modeling techniques (Sun et al. 2014; Hajek and Michalak 2013; Yeh 2012). The strength of two-stage efficiency analysis model based on for example, TOPSIS and artificial neural networks (Wanke et al. 2016), lies with not only identifying the most efficient DMUs in terms of similarity to the ideal, but also identifying the most critical contextual variables to improve efficiency. We contribute to the current state of the art by integrating TOPSIS and Neural Networks in a two-stage approach to predict the factors affecting the efficiency of diabetes treatment. We demonstrate the applicability of our proposed methodology on a dataset from diabetes health care facilities in the UK.

14.3 Methodology

The data on 213 UK health care facilities in 2012 and 211 UK health care facilities in 2013 were obtained from the Health and Social Care Information Centre (http://www.hscic.gov.uk/diabetesinpatientaudit). The TOPSIS alternatives consisted of 424 (213 in 2012 + 211 in 2013) DMUs formed by the combination of 230 unique diabetes care providers for 2 years. Inputs and outputs were those that have been commonly studied in previous works, *Physical and Human Resources Involved*, such as the hospital beds, the nursing hours, the consultant hours, the dietician hours, the podiatrist hours, and the visit by specialist diabetes team; *Monitoring Procedures*, such as the blood glucose monitoring days; *Operational Errors*, such as the medication errors, the prescription errors, the management errors, and the insulin errors, which negatively affect efficiency levels; and *Production Measures*, such as the number of diabetic patients, the number of appropriate days, the number of good diabetes days; *Patient Perceptions on Service Quality*, such as the percent of suitable meals, the percent of meals timing, the percent of meals choice, the percent of staff knowledge in answering questions, the percent of staff knowledge in working together, and the percent of overall satisfaction, which positively affect efficiency levels. Further information about the variables, their measurement units, and descriptive statistics are given in Appendix 1 (Tables 14.2 and 14.3).

14.3.1 TOPSIS

The TOPSIS method calculates the distance from the positive and negative ideal alternatives and ranks the DMUs that are closer to the positive ideal and farther from the negative ideal higher (Hwang and Yoon 2012). The ideal solution maximises the benefit and minimises the cost (Seçme et al. 2009).

In addition to the TOPSIS variables presented in Appendix 1, we analyse 28 contextual variables to explain the differences in efficiency. These variables of **health care demographics** are related to the **hospital demographics**, such as country, hospital site, year, provider code, and % diabetes; the **patient demographics**, such as white, emergency, % mgt diabetes, % other, and % non-medical; the diabetes type, such as % type 1, % type 2 insulin, % type 2 non-insulin, % type 2 diet only, % type other, and admitted with foot disease; the **medical routines**, such as seen by the MDT within 24 h, the foot risk assessment completed within 24 h, the foot risk assessment after 24 h only, and the foot risk assessment completed during the hospital stay; the **treatment dynamics and patient involvement**, such as the % Severe Hypo, the % Minor Hypo, the patient reported unexpected high, patient reported unexpected low, involved in treatment plan, and able to take control of diabetes care; and the **staff knowledge on diabetes**, such as staff awareness of diabetes, some all or most staff know enough about diabetes as perceived by the patient.

In line with the TOPSIS methodology, the best alternative should be closest to the positive-ideal solution, and farthest distance from the negative-ideal solution (Ertugrul and Karakasoglu 2009). A difference observed between the DEA and TOPSIS is how the efficiency ranks are produced. DEA minimises the distance from each DMU to the efficient frontier (Charnes et al. 1978; Banker et al. 1984) whereas TOPSIS applies Euclidean distance functions on normalised vectors of positive and negative criteria, weightings of which have already been defined by the decision-maker (Akbari et al. 2011). In synthesis, the major advantages of TOPSIS methods over DEA models are that: (1) decision-makers determine the criteria weightings, and; (2) convexity of data is not required.

We follow the standard procedure of a TOPSIS analysis. We develop an evaluation matrix of m alternatives and n criteria. We then normalise this matrix and calculate the weighted normalised matrix. We identify the worst and the best alternatives and then calculate the distance between the alternative m and the worst alternative as well as the distance between the alternative m and the best alternative. Afterwards, we calculate the similarity to the best alternative and the worst alternative to finally rank the alternatives according to this similarity.

14.3.2 A Neural Network Approach to Predicting Efficiency

We employ neural networks to complement our ranking analysis, to identify the factors that are the most effective in efficiencies of the diabetes care units. Although there is a trade-off between prediction and interpretation when neural networks and sophisticated statistical methods are used (James et al. 2013; Kuhn and Johnson 2013; Kahraman et al. 2013), we are most interested in identifying the critical factors in predicting efficiency to inform future policies to improve diabetes care. We apply the ideas in Kuhn and Johnson (2013) on managing the trade-off between interpretability and predictability. When the objective is high accuracy in prediction, the highest performing models should be preferred, even if they are not easy to interpret.

Diabetes treatment efficiency should be predicted to elicit what makes high performers more efficient since low efficiency might be associated with death or acute complications. A predictive model can help preventing or mitigating such undesired outcomes. That is why we investigate contextual variables in predicting diabetes care efficiency. The recent literature has focused on two stage models where the efficiency scores are estimated in the first stage and then followed by a predictive model in the second stage to identify the most influential factors to efficiency (Azadeh et al. 2014).

Artificial Neural Networks are composed of a set of neurons connected to each other for integrated computations (Claveria and Torra 2014). Each neuron takes as input a linear combination of explanatory variables and using a nonlinear transformation function, calculates an output (Torgo 2011). The weights of inputs are updated by error propagation from the output, which is reached through computations in hidden layers with nonlinear transformation functions (Kuhn and Johnson 2013). The number of hidden layers and the number of neurons in each layer are among the design characteristics of neural networks (Ledolter 2013; Feng and Zhang 2014).

It is not possible to estimate the parameters of a neural network from the data. The changes on these parameters are referred to as the process of ‘tuning’ and there is no analytical formula to estimate the optimum parameter values (Kuhn and Johnson 2013), which makes design of neural networks as much an art as science. It is common to use cross-validation to tune the parameters and avoid overfitting, which is learning the data set very well but not being able to generalise to unobserved data points (Kuhn and Johnson 2013). Accuracy of the model is informative in the tuning process; when it begins to increase slowly with changing parameters, overfitting starts to occur.

14.4 Results and Discussion

We performed the analyses reported in this chapter using the `topsis` package for efficiency computations, `nnet` package for neural network computations and `caret` package for extracting relative importance of contextual variables in `R` Language.

The efficiency levels calculated for 230 UK health care facilities from 2012 to 2013 using the TOPSIS approach and considering different grouping criteria: year, country, and health care facility type as well as ethnic origin of the patient are given in Figs. 14.1 and 14.2; additionally, the complete efficiency ranking is presented in Appendix 2 (Table 14.4). Readers should note that data are treated considering a meta-frontier approach, so that abnormal variations from 1 year to another are smoothed towards a common production frontier that encompasses all years and hospitals.

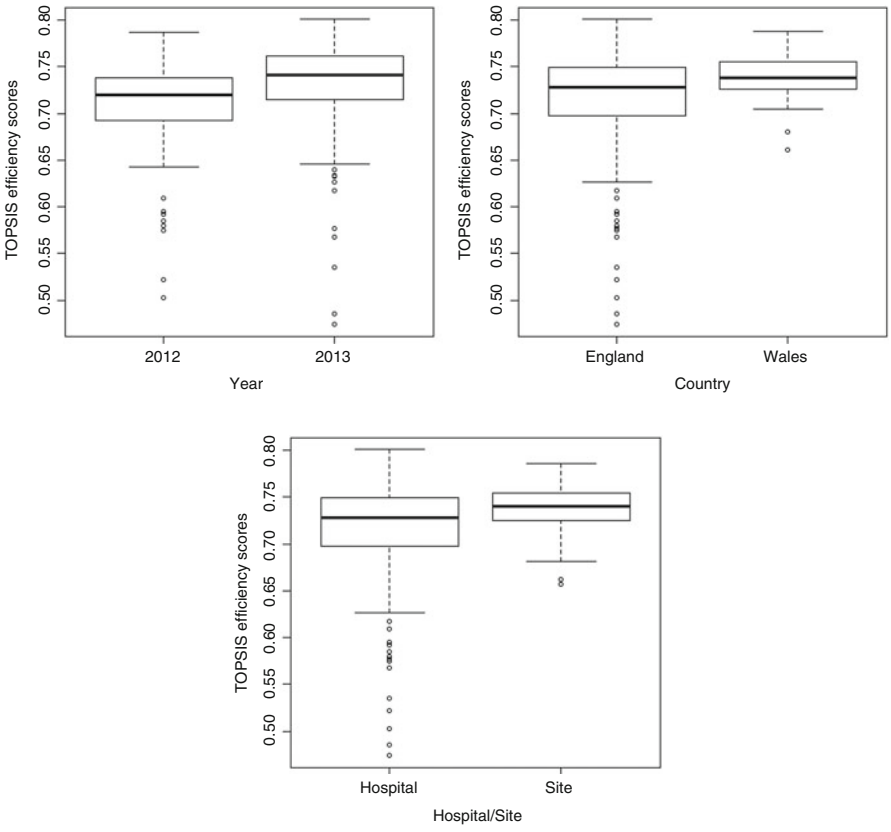

Fig. 14.1 Efficiency levels grouped by year, country and hospital/site

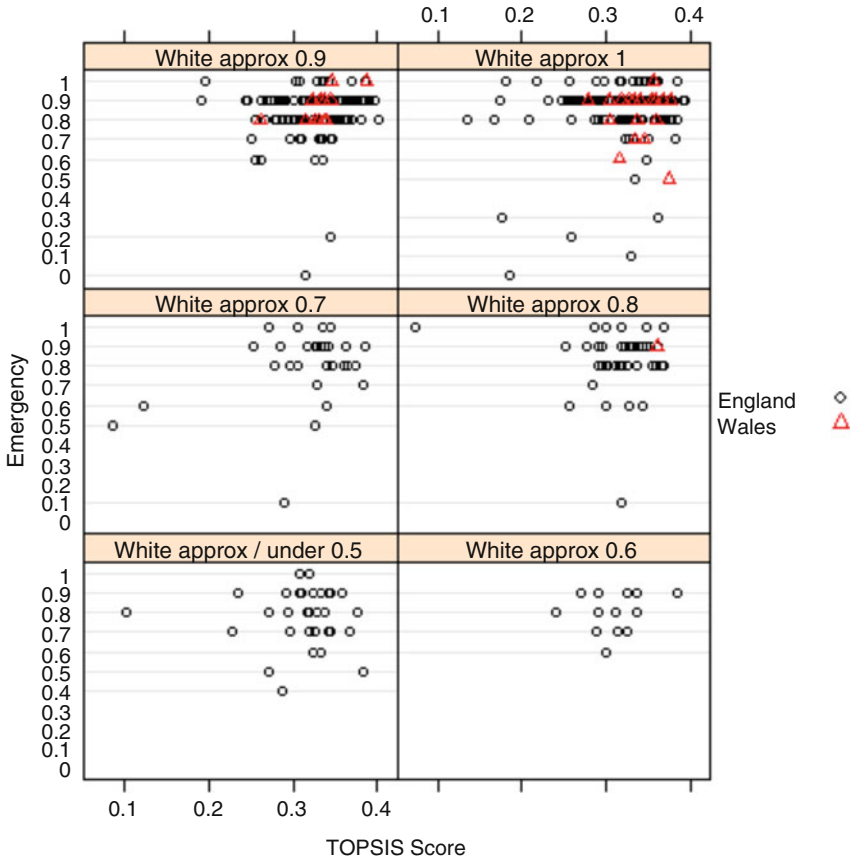

Fig. 14.2 Efficiency levels per country grouped by white and emergency values

We performed a robustness analysis to compare TOPSIS scores with those computed from the traditional BCC (Banker et al. 1984) and CCR (Charnes et al. 1978) models (cf. Fig. 14.3). We need to assess if the TOPSIS method has a higher discriminatory power and if the efficiency scores have high contrast, i.e. are less symmetrical around the mean. High contrasts are necessary for the following predictive analysis.

The mean overall efficiency scores in the TOPSIS method is 0.72, whereas the traditional BCC and CCR models presented mean values of 0.98 and 0.79, respectively. This result suggests that the discriminatory power of the TOPSIS method is higher than those observed in traditional models, because their scores are less biased towards one. TOPSIS responds better than traditional DEA models in light of numerous inputs and outputs, an effect usually known as the curse of dimensionality. This effect cause DEA scores to be strongly biased towards one. The impact of TOPSIS efficiency modelling can be also found in other statistical

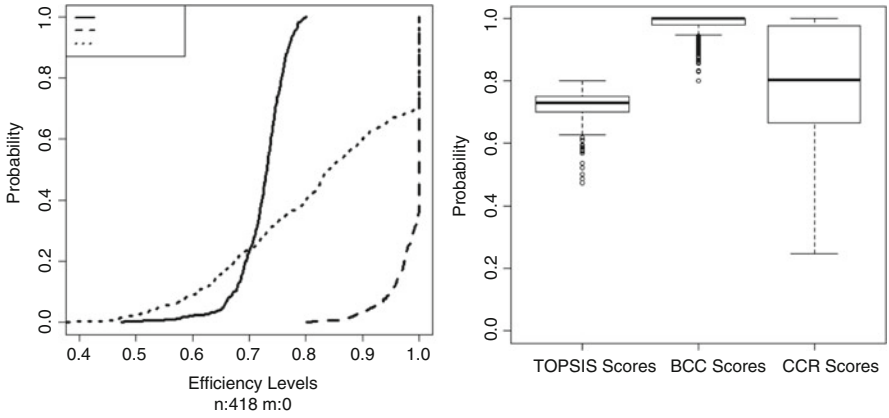

Fig. 14.3 Robustness analysis

properties derived from the frequency distribution of efficiency estimates in both models. Skewness is closer to zero (-1.767 [TOPSIS] against -2.680 [BCC]), suggesting that, in the TOPSIS method, efficiency scores are not symmetrical around the mean, favouring the use of different predictive models in general. Besides, due to the bias towards one, the Spearman rank correlations between efficiency scores derived from traditional DEA models and the TOPSIS method were found to be low (0.26 for the BCC model).

We ran the global separability (Daraio et al. 2010) test for contextual variables and found that the test statistic for TOPSIS (0.064) indicated good separability. Values of this statistic approaching 1 indicate bias (cf. Fig. 14.3). Global separability, therefore, appears to be consistent with the use of TOPSIS on the sample data. This suggests that the contextual variables considered here affect only the distribution of efficiencies and not the attainable input/output combinations.

We also performed a robustness analysis prior to running the neural network analysis. In the robustness analysis, we regressed TOPSIS efficiency scores against the contextual variables (Table 14.1) using Tobit regression (see Moyo (2012) for further details on Tobit regression). The results in Table 14.1 indicate a number of contextual variables that impact efficiency levels in the treatment of diabetes in the UK health care facilities. Efficiency appears to be increasing over the course of time and is higher in sites receiving a higher proportion of emergency patients. Efficiency is also higher in Wales, when the staff is aware of diabetes, and when patients present type-2 non-insulin diabetes. Efficiency is lower when patients present other diabetes types and minor hypoglycemic conditions.

Then, we performed a neural network analysis on TOPSIS efficiency scores where we assigned contextual variables presented in Section 3 as predictors. We followed the steps suggested in Faraway (2016) and Kuhn and Johnson (2013).

Figure 14.4 shows a common pattern within the cross-validation methods. The RMSE is lower for lower number of hidden layers and increases as the number

Table 14.1 Tobit regression results

Coefficients	Estimate	Std. error	z value	$P(>\|z\|)$
(Intercept)	−88.104	31.502	−2.797	0.005∗
Country	0.014	0.008	1.68	0.093.
Hospital/Site	0.017	0.008	2.267	0.023∗
Year	0.044	0.016	2.813	0.005∗
White	0.022	0.015	1.474	0.14
Emergency	0.04	0.016	2.478	0.013∗
% $Mgt_Diabetes$	−0.014	0.053	−0.257	0.797
% Other	−0.065	0.021	−3.025	0.002∗
% $Type_1$	0.068	0.113	0.599	0.549
% $Type_2$ insulin	0.134	0.1	1.335	0.182
% $Type_2$ non insulin	0.179	0.1	1.788	0.074.
% $Type_2$ diet only	0.102	0.103	0.994	0.32
Admitted with ft dis.	0.005	0.05	0.098	0.922
Seen by the MDT with 24h	0.001	0.007	0.217	0.828
Ft risk ass. comp. with 24h	0.009	0.009	1.042	0.297
Ft risk ass. after 24h only	0.042	0.026	1.635	0.102
% Severe hypo	0.038	0.044	0.872	0.383
% Minor hypo	−0.07	0.034	−2.084	0.037∗
Patient rep. unexpected high	0.024	0.019	1.235	0.217
Patient rep. unexpected low	0.057	0.023	2.475	0.013∗
Involved in treatment plan	−0.005	0.017	−0.317	0.751
Able take ctrl. diabetes care	−0.01	0.019	−0.561	0.575
Staff awareness of diabetes	0.036	0.019	1.905	0.057.
Staff knowledge of diabetes	0.031	0.02	1.52	0.129
Log(scale)	−3.167	0.035	−91.567	0

Signif. codes: "*" sig. at 0.05 "." sig. at 0.10; Scale: 0.04213; Gaussian distribution; Number of Newton-Raphson Iterations: 4; Log-likelihood: 730.6 on 25 Df; Wald-statistic: 730.6 on 25 Df

of hidden layers increase. The most accurate neural network was obtained (RMSE = 0.041) using the repeated 10-fold cross-validation technique, for fifteen hidden layers and a decay rate of 0.008 (Fig. 14.4). This suggests a mean error rate of 4.14% considering that the efficiency scale ranges from 0 to 100%. The number of folds used in all cross-validation methods was 25, except for the 10-fold cross validation and the repeated 10-fold cross-validation.

Figure 14.5 presents the relative importance of contextual variables. The variables related to the hospital and patient demographics are the top six predictors of efficiency scores of the UK diabetes care facilities. Variables related to the treatment dynamics and patient involvement and to the staff knowledge on diabetes appears to be the least important efficiency predictors. Medical routines present an intermediate importance.

Figure 14.6 presents the one-dimensional sensitivity analysis on the TOPSIS efficiency estimates for the best neural network model as described in Cortez and

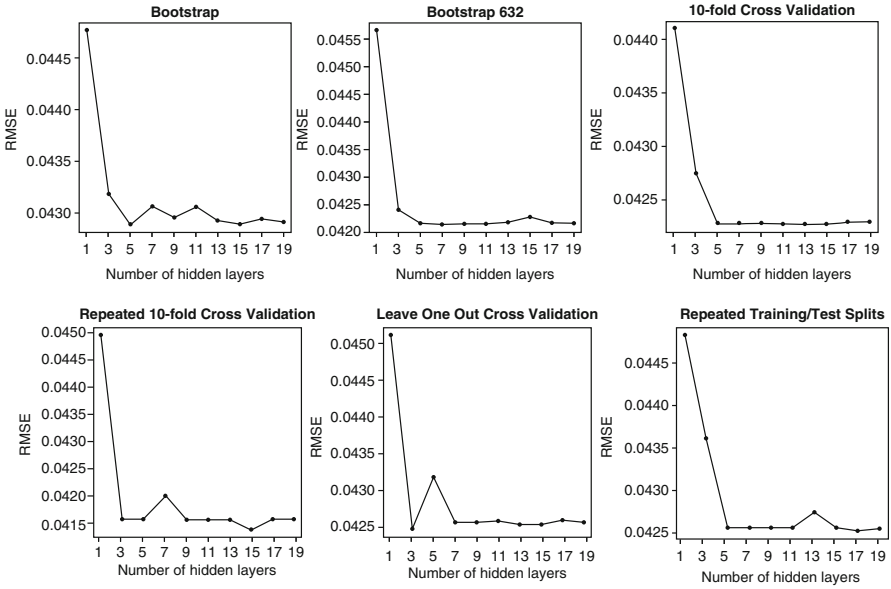

Fig. 14.4 Cross-validated performance profiles over different tuning parameter values

Embrechts (2013). We standardised the contextual variables before performing the marginal analysis, as suggested in Faraway (2016). Vertical axis represents the efficiency scores in the sensitivity analysis for the predicted relationship between the outcome (efficiency) and the standardised predictors (contextual variables) while keeping all other standardised predictors at their mean value (zero due to standardisation). Both the magnitude of the scales on the vertical axis and their variation intervals are related to the maximum and minimum TOPSIS scores. The similitude of the vertical axis for different predictors or contextual variables is related to their standardisation. However, their marginal variation with respect to the unit, holding all other predictors at their mean value, should be interpreted in terms of the magnitude of the values of the horizontal axis, which tends to vary substantially, depending upon the predictor.

Results confirmed the signs found in Tobit regression results, irrespective of their significance levels and relative importance, as measured by the regression coefficients. They suggest that additional variables on the patient demographics and hospital demographics should be measured and incorporated into future studies since these are the most relevant predictors of efficiency when compared to treatment dynamics and medical routines. Variables related to gender, age, educational level, professional background, as well as socio-economic indicators such as average income and human development level should be taken into account in order to clearly map which regions should be target by authorities for improvement plans.

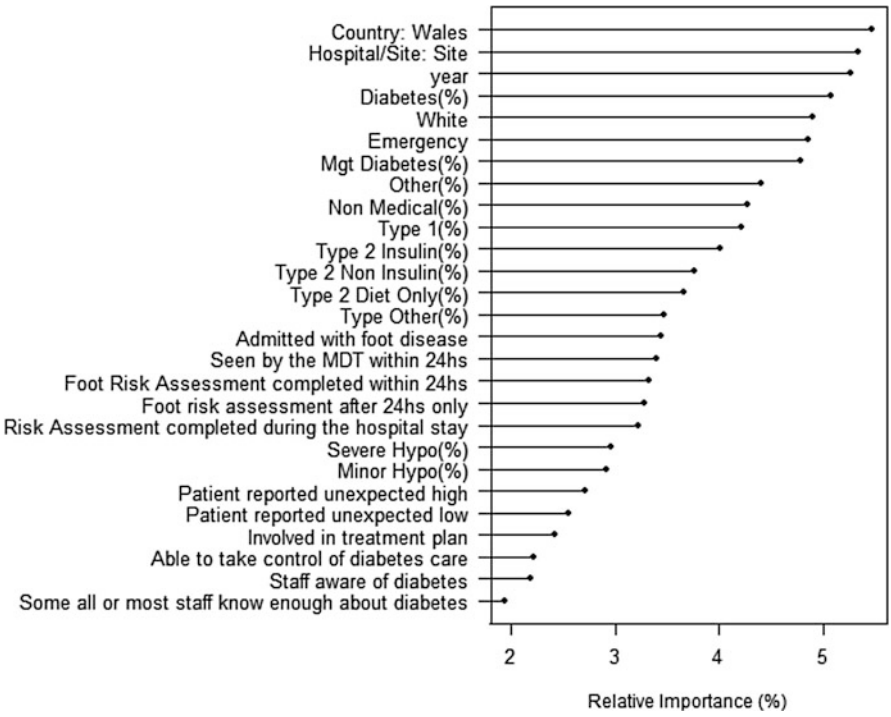

Fig. 14.5 Relative importance of contextual variables

The fact that treatment dynamics and medical routines presented intermediate to lower importance as efficiency predictors suggest that these variables seem to be quite standardized across different demographic segments, such as country, type of health care facility, and over the course of time.

14.5 Conclusion

This chapter presents an analysis of the efficiency of diabetes treatment in the UK health care facilities using TOPSIS and neural networks. TOPSIS enables a ranking of the efficiency of the health care facilities analyzed and, based on such ranking, a high variation in efficiency can be discerned, as presented in Appendices 1 and 2. The efficiency of diabetes care providers is affected by a number of contextual variables, among which, the most important are the prevalence of diabetes, the ethnic background of the patient, whether the type of admission was emergency or not, and whether the patient was involved in the management of their diabetes. As is evidenced by TOPSIS efficiency scores and the neural network analysis, contextual variables have profound implications for diabetes management because

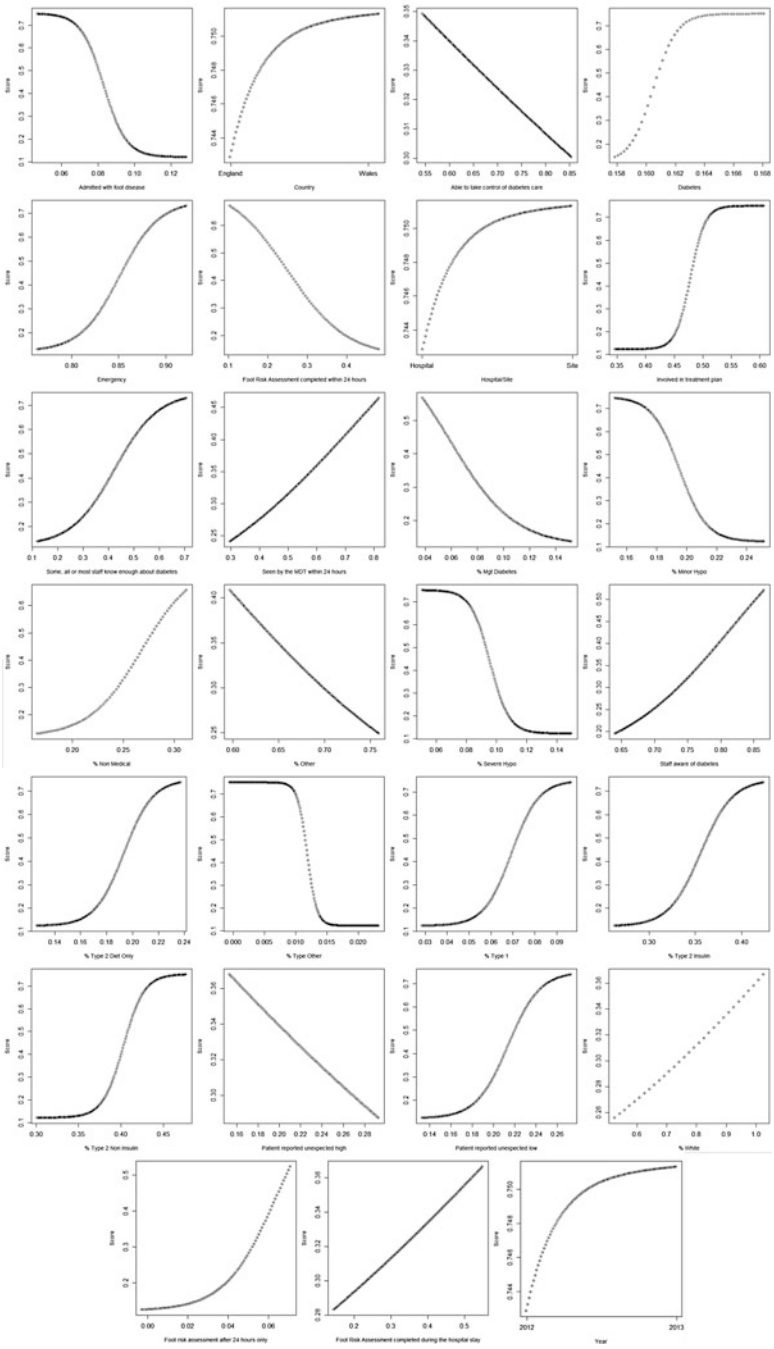

Fig. 14.6 Neural network sensitivity analysis for the efficiency scores

they can affect the quality of life of the patient and the cost of diabetes care services. We provide a rational framework for policy makers to rank the efficiency of diabetes care facilities and also highlight the most important contextual variables that impact on the efficiency as issues of interest for future policies. For example, it is known that lifestyle and diet are important determinants of diabetes (Nguyen et al. 2016). To help manage and reduce the prevalence, future policies could focus on encouraging active lifestyle choices and healthy diets targeted differently at people with diabetes and the general population as well as people from different ethnic backgrounds. Moreover, as our sensitivity analysis confirms, the patients' perspective in managing the disease is also relevant; so, patient engagement policies for raising awareness of not only patients but also their families could be developed in order to assure high patient understanding of the disease, which would impact directly on the management of the disease. As is evidenced by the results, there is no single solution to addressing diabetes care; therefore, to increase the effectiveness of care services, different models should be developed and adopted depending on the context (Amado and Dyson 2009). Moreover, definitions of measures may change over time, which should be considered when making inferences from longitudinal data (McCullough et al. 2015).

The limitations of our work should also be acknowledged. We have worked with open data published by the NHS UK; therefore, our results are bound by the limitations associated with data collection. There may be some contextual variables which are not included in the analysis simply because data on them was not collected during the audit. The data is cross-sectional and static compared to the dynamic nature of service provision. Apart from that, we believe our methodology is objective and robust to understand the efficiency of diabetes care providers in the context of England and Wales.

Future research could easily follow up with new data published by the audit to compare the progress of efficiencies of different care providers and to determine whether the most relevant contextual variables change over time. Another venue for research could be to assess the efficiencies of various diabetes care and management policies over time. Such analyses could address the dynamic nature of efficiency evaluation. If comparable data is available from another country, then an international benchmarking study could identify the best practices as well as differences in diabetes care provision.

Appendix 1: Contextual Variables and Their Descriptives

Table 14.2 Contextual variables and their measurement

Variable name	Variable type	Variable name	Variable type
Diabetic patients (BA)	Integer	Management errors	Percentage 0%
Prevalence of diabetes	Percentage 0%	Insulin errors	Percentage 0%
Percentage White	Percentage 0%	Admitted with foot disease	Percentage 0%
Percentage emergency	Percentage 0%	Seen by the MDT within 24 h	Percentage 0%
Management of diabetes	Percentage 0%	Foot Risk Assessment completed within 24 h	Percentage 0%
Other medical	Percentage 0%	Foot risk assessment after 24 h only	Percentage 0%
Non-medical	Percentage 0%	Foot risk assessment completed during the hospital stay	Percentage 0%
Diabetes Type 1	Percentage 0%	% Severe hypo	Percentage 0%
Diabetes Type 2—Insulin	Percentage 0%	% Minor hypo	Percentage 0%
Diabetes Type 2—Non insulin	Percentage 0%	Meals suitable	Percentage 0%
Diabetes Type 2—Diet only	Percentage 0%	Meals timing	Percentage 0%
Diabetes Type Other	Percentage 0%	Meals Choice	Percentage 0%
Nursing hours	Decimal(4,2)	Staff knowledge—answer Qs	Percentage 0%
Consultant hours	Decimal(4,2)	Staff Knowledge—Emotional support	Percentage 0%
Dietician hours	Decimal(4,2)	Staff Knowledge—Work together	Percentage 0%
Podiatrist hours	Decimal(4,2)	Overall Satisfaction	Percentage 0%
Pharmacist hours	Decimal(4,2)	Patient reported unexpected high	Percentage 0%
Blood glucose monitoring days	Decimal(4,2)	Patient reported unexpected low	Percentage 0%
Appropriate days	Decimal(4,2)	Able to take control of diabetes care	Percentage 0%
Good diabetes days	Decimal(4,2)	Staff aware of diabetes	Percentage 0%
Visit by specialist diabetes team	Percentage 0%	All or most staff know enough about diabetes	Percentage 0%
Medication errors	Percentage 0%	Renal replacement therapy	Percentage 0%
Prescription errors	Percentage 0%	Number of returned Patient Experience forms (PE)	Integer

Table 14.3 Descriptive statistics for the TOPSIS criteria and the contextual variables

	Min	Q1	Q2	Mean	Q3	Max
TOPSIS criteria						
Hospital beds	46	265.5	375.5	419.7	530	1327
Nursing hours	0	1.055	1.725	2.02	2.5	16.5
Consultant hours	0	0.443	0.7	0.845	1.035	4.8
Dietician hours	0	0.2	0.4	0.571	0.73	4.86
Podiatrist hours	0	0.143	0.34	0.556	0.708	5.67
Blood glucose monit'g days	0	0	5.8	3.4	6.775	7
Medication errors	0	0.283	0.38	0.375	0.463	0.933
Prescription errors	0	0.136	0.222	0.224	0.308	0.6
Management errors	0	0.154	0.219	0.226	0.289	0.867
Insulin errors	0	0.134	0.2	0.205	0.262	0.55
Diabetic patients	5	39.25	59	65.62	85	234
Appropriate days	2.9	6.4	6.7	6.528	6.8	7
Good diabetes days	2.4	4.125	5.6	5.288	6.5	7
Visit by specialist team	0	0.24	0.307	0.337	0.4	0.967
Meals suitable	0.542	0.803	0.873	0.863	0.928	1
Meals timing	0.233	0.65	0.733	0.73	0.819	1
Meals choice	0	0.566	0.656	0.658	0.757	1
Staff: answer Qs	0.186	0.721	0.81	0.799	0.881	1
Staff: emotional support	0.388	0.792	0.876	0.856	0.937	1
Staff: work together	0	0.357	0.449	0.447	0.53	1
Overall satisfaction	0.458	0.798	0.875	0.86	0.929	1
Contextual and business-related characteristics						
%Diabetes	0.013	0.133	0.156	0.16	0.183	0.349
%Mgt diabetes	0	0.042	0.071	0.078	0.111	0.265
%Other	0.217	0.619	0.672	0.673	0.733	1
%Non medical	0	0.19	0.244	0.249	0.307	0.733
%Type 1	0	0.036	0.061	0.065	0.088	0.467
%Type 2 insulin	0.059	0.28	0.336	0.343	0.404	0.867
%Type 2 non insulin	0	0.333	0.394	0.394	0.462	0.75
%Type 2 diet only	0	0.136	0.179	0.185	0.224	0.667
%Type other	0	0	0	0.012	0.02	0.2
%Severe hypo	0	0.061	0.095	0.1	0.133	0.306
%Minor hypo	0	0.155	0.197	0.202	0.241	0.5
White	0	0.811	0.92	0.859	0.971	1
Emergency	0	0.799	0.875	0.836	0.919	1
Admitted with foot disease	0	0.053	0.083	0.088	0.118	0.283
Seen by MDT within 24 h	0	0.333	0.5	0.523	0.778	1
Foot risk ass't in 24 h	0	0.125	0.224	0.306	0.423	1
Foot risk ass't after 24 h	0	0	0.029	0.054	0.071	0.676
Foot risk ass't in hospital	0	0.167	0.294	0.36	0.512	1
Patient unexpected high	0	0.151	0.21	0.223	0.283	1
Patient unexpected low	0	0.132	0.204	0.203	0.264	0.601

(continued)

Table 14.3 (continued)

	Min	Q1	Q2	Mean	Q3	Max
Involved in treatment plan	0	0.361	0.461	0.469	0.578	1
Can take control of diabetes	0	0.551	0.711	0.69	0.841	1
Staff aware of diabetes	0.291	0.674	0.754	0.753	0.852	1
Staff knowledge of diabetes	0	0.017	0.275	0.362	0.694	1
		(%)				
Country	England	92				
	Wales	8				
Hospital site	Hospital	92				
	Site	8				
Year	2012	50				
	2013	50				

Appendix 2: Efficiency Ranking

Table 14.4 Efficiency rankings of DMUs

Year	Provider code	TOPSIS score	Ranking	Year	Provider code	TOPSIS score	Ranking
2013	RF4QH	0.474	1	2012	RJC02	0.729	210
2013	RTH02	0.485	2	2013	7A4BV	0.73	211
2012	RV831	0.502	3	2013	REF12	0.73	212
2012	RTH02	0.522	4	2012	RA7	0.73	213
2013	RCD01	0.535	5	2013	RWF03	0.731	214
2013	RD304	0.567	6	2013	7A3C7	0.731	215
2012	RCD01	0.575	7	2012	RTX02	0.731	216
2013	RWP31	0.577	8	2012	RXL01	0.731	217
2012	RNLAY	0.58	9	2012	RLNGL	0.732	218
2012	RWP31	0.585	10	2013	RWFTW	0.732	219
2012	RD304	0.592	11	2012	7A3B7	0.732	220
2012	RTH05	0.595	12	2012	RXPCP	0.732	221
2012	RWF03	0.609	13	2013	RXH09	0.732	222
2013	RP5BA	0.618	14	2012	RJ611	0.732	223
2013	RVLC9	0.626	15	2012	RWA	0.732	224
2013	RA901	0.632	16	2012	RFSDA	0.732	225
2013	RV831	0.633	17	2012	RN325	0.733	226
2013	RQM22	0.64	18	2012	RAL01	0.733	227
2012	RFRPA	0.642	19	2013	RJR05	0.733	228
2013	RL410	0.646	20	2013	RAX01	0.733	229
2012	RP5BA	0.646	21	2012	RBD01	0.733	230

(continued)

Table 14.4 (continued)

Year	Provider code	TOPSIS score	Ranking	Year	Provider code	TOPSIS score	Ranking
2012	RTE01	0.65	22	2012	RWP50	0.733	231
2012	RWH20	0.652	23	2013	RQ6	0.733	232
2012	RXK01	0.652	24	2012	RXWAS	0.733	233
2013	RTXBU	0.655	25	2012	RGP75	0.733	234
2012	RA901	0.655	26	2013	7A1A4	0.734	235
2013	RW602	0.656	27	2013	RD7	0.734	236
2012	RKB03	0.656	28	2013	RJC02	0.734	237
2013	RVY	0.657	29	2013	7A2AL	0.734	238
2013	RTR45	0.657	30	2012	RXPBA	0.735	239
2013	RTFED	0.658	31	2012	RBA11	0.735	240
2013	RTFDR	0.659	32	2012	RXF10	0.735	241
2012	RK5HP	0.659	33	2013	RXWAT	0.735	242
2013	7A2AJ	0.661	34	2013	RXN02	0.736	243
2013	RW601	0.661	35	2012	RN707	0.736	244
2012	RTR45	0.661	36	2012	RDU01	0.736	245
2013	RTD	0.662	37	2013	RW603	0.736	246
2013	R1F01	0.662	38	2013	RA7	0.736	247
2013	RFRPA	0.663	39	2013	RKEQ4	0.736	248
2013	RCC25	0.666	40	2013	RBD01	0.737	249
2012	RXH09	0.666	41	2012	RYR16	0.737	250
2013	RLQ01	0.668	42	2012	7A4BV	0.737	251
2012	RWY02	0.669	43	2012	RJR05	0.738	252
2012	RJ121	0.67	44	2013	RD899	0.738	253
2012	RFW01	0.67	45	2012	7A3C4	0.738	254
2012	RYJ01	0.671	46	2013	R1H12	0.738	255
2012	RJ201	0.671	47	2013	7A3C4	0.738	256
2013	RJL32	0.671	48	2012	RW3	0.738	257
2012	RTFFS	0.672	49	2012	RBL14	0.738	258
2012	RL410	0.677	50	2012	RD130	0.738	259
2012	R1F01	0.678	51	2012	RF4DG	0.738	260
2012	RX1RA	0.678	52	2013	RMC01	0.739	261
2013	RP5DR	0.678	53	2013	RLT99	0.739	262
2012	RAE05	0.679	54	2012	RMP01	0.739	263
2012	RBZ12	0.679	55	2012	RH801	0.739	264
2013	RWDLP	0.679	56	2012	7A1AU	0.74	265
2012	RJD01	0.68	57	2012	RVJ	0.74	266
2012	RNLBX	0.68	58	2012	RE9GA	0.74	267
2012	7A2AJ	0.68	59	2013	RAE	0.74	268
2013	RE9GA	0.681	60	2013	RVJ	0.741	269
2012	RBN	0.681	61	2013	RYR18	0.741	270
2013	RVWAA	0.681	62	2012	RRK02	0.741	271

(continued)

Table 14.4 (continued)

Year	Provider code	TOPSIS score	Ranking	Year	Provider code	TOPSIS score	Ranking
2012	RVR05	0.685	63	2013	7A3B7	0.741	272
2013	RKB03	0.685	64	2013	RYJ03	0.741	273
2012	RVWAA	0.685	65	2013	RVR50	0.741	274
2013	RXK01	0.685	66	2012	RA430	0.741	275
2012	RNJM0	0.686	67	2012	RXK02	0.741	276
2013	RTE03	0.686	68	2013	RXPDA	0.741	277
2012	RVWAE	0.686	69	2013	RTGFG	0.742	278
2012	RXQ50	0.687	70	2013	RF4DG	0.742	279
2012	RXPDA	0.687	71	2012	RNHB1	0.742	280
2012	RXN01	0.687	72	2013	RNQ51	0.743	281
2012	RDZ20	0.687	73	2013	RDZ20	0.743	282
2012	RWFTW	0.687	74	2013	RN707	0.743	283
2012	RJL32	0.688	75	2012	REF12	0.743	284
2012	RJ701	0.688	76	2013	RHQ1	0.744	285
2012	RJ122	0.689	77	2012	RC971	0.744	286
2013	RBZ12	0.689	78	2013	RQ8L0	0.744	287
2012	RTE03	0.69	79	2012	RAJ01	0.744	288
2012	RNJ83	0.69	80	2012	RWG	0.744	289
2012	RAS01	0.69	81	2012	RM301	0.744	290
2012	RW3TR	0.69	82	2013	RBF03	0.745	291
2012	RR7EN	0.691	83	2013	RWDDA	0.745	292
2012	RTFED	0.691	84	2013	RXK02	0.745	293
2013	RJ201	0.691	85	2013	RYJ02	0.745	294
2013	RTE01	0.691	86	2013	RH801	0.745	295
2012	RCF22	0.692	87	2013	RA430	0.745	296
2012	RW603	0.692	88	2013	RNS01	0.745	297
2013	RC399	0.692	89	2012	7A4C1	0.745	298
2012	RTXBU	0.692	90	2013	RVLC7	0.746	299
2012	RXH01	0.693	91	2012	7A2AG	0.746	300
2013	RYJ01	0.694	92	2013	7A2BL	0.746	301
2012	RC399	0.695	93	2013	RWG	0.747	302
2012	RW601	0.695	94	2013	RWP01	0.747	303
2013	RX1CC	0.696	95	2012	RTK	0.747	304
2012	RVLC9	0.696	96	2012	RDE	0.748	305
2013	RWH20	0.697	97	2013	RWJ09	0.748	306
2012	RVV01	0.697	98	2012	RCX70	0.748	307
2013	RJD01	0.697	99	2012	RTRAT	0.749	308
2012	RVLC7	0.697	100	2013	RDE	0.749	309
2012	RN506	0.697	101	2013	RTP04	0.749	310
2013	RVV01	0.698	102	2013	RYQ50	0.749	311
2013	RWH01	0.699	103	2013	RXR20	0.749	312

(continued)

Table 14.4 (continued)

Year	Provider code	TOPSIS score	Ranking	Year	Provider code	TOPSIS score	Ranking
2012	RX1CC	0.699	104	2013	RXC01	0.749	313
2012	RJF02	0.7	105	2012	RYR18	0.749	314
2012	RRV03	0.7	106	2012	RHQNG	0.75	315
2012	RF4QH	0.701	107	2013	RTRAT	0.75	316
2012	RP5DR	0.701	108	2013	RXF05	0.75	317
2012	RLQ01	0.701	109	2012	RHM01	0.751	318
2013	RNLBX	0.701	110	2013	RA301	0.752	319
2013	RNLAY	0.702	111	2012	RYQ30	0.752	320
2013	RTH05	0.702	112	2013	RK950	0.752	321
2013	RVR05	0.703	113	2013	RN325	0.752	322
2012	RGQ01	0.704	114	2013	RWEAE	0.753	323
2012	RWDLP	0.704	115	2012	RVY	0.753	324
2013	7A1A1	0.705	116	2013	RM102	0.754	325
2012	7A2BL	0.705	117	2012	RHW01	0.754	326
2012	RD7	0.706	118	2012	RNA01	0.754	327
2012	RTH08	0.706	119	2012	RRF	0.754	328
2013	RBT20	0.706	120	2012	RXWAT	0.755	329
2012	RKEQ4	0.707	121	2013	RJF02	0.755	330
2012	RNJ12	0.707	122	2013	RR8	0.755	331
2012	RXC01	0.708	123	2013	RLNGL	0.755	332
2012	RTFDR	0.708	124	2012	RA299	0.755	333
2012	RTD	0.709	125	2013	7A5B1	0.756	334
2012	RHU03	0.709	126	2013	RWEAK	0.756	335
2013	RGT01	0.709	127	2013	RC110	0.756	336
2012	RGT01	0.709	128	2012	7A5B1	0.756	337
2013	RQX99	0.71	129	2012	RGR01	0.756	338
2012	RAE01	0.711	130	2013	RWP50	0.756	339
2012	RQQ31	0.711	131	2012	RXQ02	0.757	340
2012	RVR50	0.711	132	2012	7A5B3	0.758	341
2012	REM21	0.711	133	2013	RV820	0.758	342
2012	RM102	0.711	134	2013	RJN71	0.758	343
2012	RTP04	0.711	135	2013	RN506	0.759	344
2012	RQ6	0.712	136	2013	RQQ31	0.759	345
2012	RM202	0.712	137	2012	7A6AM	0.759	346
2012	RQ8L0	0.713	138	2013	RXPCP	0.76	347
2013	RXH01	0.713	139	2013	RTH08	0.76	348
2012	RXR10	0.714	140	2013	RBK02	0.76	349
2013	RXWAS	0.714	141	2013	RFF99	0.761	350
2012	RXR20	0.714	142	2012	RWJ09	0.761	351
2012	7A1A1	0.715	143	2012	RWWWH	0.761	352
2013	RAS01	0.715	144	2012	7A2AL	0.761	353

(continued)

Table 14.4 (continued)

Year	Provider code	TOPSIS score	Ranking	Year	Provider code	TOPSIS score	Ranking
2013	RXN01	0.715	145	2013	7A6AR	0.761	354
2013	RWDLA	0.716	146	2013	RXPBA	0.761	355
2013	RX1RA	0.716	147	2013	RCX70	0.761	356
2013	RFW01	0.716	148	2012	RBK02	0.761	357
2013	RWY02	0.716	149	2013	RTX02	0.761	358
2012	7A6BF	0.717	150	2012	RR105	0.761	359
2012	RJZ01	0.717	151	2013	RQWG0	0.762	360
2013	REM21	0.717	152	2013	RXC02	0.763	361
2013	RR7EN	0.717	153	2012	RNZ00	0.763	362
2013	RCF22	0.718	154	2013	RBL14	0.763	363
2012	RC110	0.718	155	2013	RBA11	0.763	364
2012	RQX99	0.718	156	2013	RRF	0.763	365
2013	RJ701	0.718	157	2013	RR10	0.764	366
2012	RYQ50	0.718	158	2013	RW3	0.764	367
2012	RKB01	0.718	159	2013	RYR16	0.764	368
2012	RWDDA	0.719	160	2012	RPA02	0.764	369
2012	RK950	0.719	161	2012	RJE01	0.765	370
2012	RBF03	0.719	162	2012	RNS01	0.766	371
2012	RD899	0.719	163	2013	RM202	0.766	372
2013	RBN	0.719	164	2013	RKB01	0.766	373
2013	RTFFS	0.719	165	2012	RYJ03	0.768	374
2012	RWEAE	0.719	166	2013	7A2AG	0.768	375
2012	7A1A4	0.719	167	2013	RR105	0.769	376
2012	RBT20	0.72	168	2013	RXQ	0.769	377
2012	RVL01	0.72	169	2013	RWEAA	0.769	378
2012	RWDLA	0.72	170	2013	RGN80	0.769	379
2012	RTGFG	0.721	171	2013	RA299	0.769	380
2013	RWA	0.721	172	2013	RDDH0	0.77	381
2012	RJN71	0.721	173	2013	RHQNG	0.77	382
2013	RFSDA	0.722	174	2013	RTK	0.771	383
2012	RWY01	0.722	175	2013	RGP75	0.771	384
2012	RJL30	0.723	176	2013	RJL30	0.772	385
2013	RD130	0.723	177	2013	RCB55	0.772	386
2012	7A6AR	0.723	178	2013	RXL01	0.773	387
2012	RV820	0.723	179	2013	RAJ01	0.773	388
2012	RW602	0.723	180	2012	RGN80	0.774	389
2013	RRV03	0.723	181	2013	RRK02	0.775	390
2012	RWWHG	0.724	182	2013	RVWAE	0.775	391
2012	RXF05	0.724	183	2013	RPA02	0.775	392
2013	RJ611	0.724	184	2013	R1HNH	0.775	393
2013	RW3TR	0.725	185	2013	7A6BF	0.776	394

(continued)

Table 14.4 (continued)

Year	Provider code	TOPSIS score	Ranking	Year	Provider code	TOPSIS score	Ranking
2013	RWY01	0.725	186	2013	RMP01	0.777	395
2012	RHQ1	0.725	187	2012	RFF99	0.778	396
2012	RR10	0.725	188	2013	RVVKC	0.778	397
2012	RDDH0	0.725	189	2013	7A6AM	0.779	398
2013	RAL01	0.725	190	2013	RWWWH	0.779	399
2012	RA301	0.726	191	2013	RHM01	0.781	400
2012	7A3C7	0.726	192	2013	RNZ00	0.782	401
2012	RXC02	0.726	193	2012	RVV09	0.783	402
2012	RYJ02	0.726	194	2013	RJ1	0.783	403
2012	RXQ01	0.726	195	2013	R1H83	0.784	404
2012	RWP01	0.726	196	2013	RVV09	0.785	405
2012	RWEAK	0.727	197	2013	RYQ30	0.785	406
2013	RXF10	0.727	198	2013	RJE	0.785	407
2012	RVVKC	0.727	199	2013	R1HKH	0.785	408
2012	RWEAA	0.727	200	2013	RGR01	0.786	409
2013	RC971	0.728	201	2013	RHW01	0.786	410
2012	RLT99	0.728	202	2012	RN101	0.787	411
2012	RGCKH	0.728	203	2013	7A5B3	0.788	412
2013	RJZ01	0.728	204	2013	RDU01	0.791	413
2013	7A4C1	0.729	205	2013	RGQ01	0.792	414
2012	RXN02	0.729	206	2013	RHU03	0.795	415
2012	RNQ51	0.729	207	2013	RN101	0.797	416
2012	RWH01	0.729	208	2013	RNA01	0.798	417
2013	RWWHG	0.729	209	2013	RM301	0.801	418

References

Afful-Dadzie E, Nabareseh S, Oplatková Z, Klímek P (2016) Model for assessing quality of online health information: a fuzzy vikor based method. J Multi-Criteria Decis Anal 23(1–2), 49–62. doi:10.1002/mcda.1558. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84959914493&partnerID=40&md5=e13dcc3d60a793774544deee6a187012

Akbari M, Shaverdi M, Fallah Tafti S (2011) Combining fuzzy MCDM with BSC approach in performance evaluation of Iranian private banking sector. Adv Fuzzy Syst 2011(148712):1–12. doi:10.1155/2011/148712. https://www.scopus.com/inward/record.uri?eid=2-s2.0-80053522025&partnerID=40&md5=4a27b86551184e41132b164117759505

Amado C, Dyson R (2009) Exploring the use of dea for formative evaluation in primary diabetes care: an application to compare english practices. J Oper Res Soc 60(11):1469–1482. doi:10.1057/jors.2008.160. https://www.scopus.com/inward/record.uri?eid=2-s2.0-70349495254&partnerID=40&md5=7963f8fd96d0993a43c74c255f1df9b0

Amaral T, Costa A (2014) Improving decision-making and management of hospital resources: an application of the promethee ii method in an emergency department. Oper Res Health Care 3(1):1–6. doi:10.1016/j.orhc.2013.10.002. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84888069808&partnerID=40&md5=c4ca28a46b5cd9d93f627ebd2cae19ab

Azadeh A, Sheikhalishahi M, Boostani A (2014) A flexible neuro-fuzzy approach for improvement of seasonal housing price estimation in uncertain and non-linear environments. S Afr J Econ 82(4):567–582. doi:10.1111/saje.12047. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84915791101&partnerID=40&md5=f185ee22079342b8f67f31d0a75afe69

Banker R, Charnes A, Cooper W (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manag Sci 30(9):1078–1092. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0021497874&partnerID=40&md5=753a401ede6e4b37701f5033a65cba03

Behzadian M, Khanmohammadi Otaghsara S, Yazdani M, Ignatius J (2012) A state-of the-art survey of topsis applications. Expert Syst Appl 39(17):13051–13069. doi:10.1016/j.eswa.2012.05.056. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84865008381&partnerID=40&md5=184a7397afa7c6fd3e8adf18866e0bc7

Bhat VN (2005) Institutional arrangements and efficiency of health care delivery systems. Eur J Health Econ 6(3):215–222. doi:10.1007/s10198-005-0294-1. http://dx.doi.org/10.1007/s10198-005-0294-1

Bouyssou D (1999) Using DEA as a tool for MCDM: some remarks. J Oper Res Soc 50(9):974–978. doi:10.1057/palgrave.jors.2600800. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0033239128&partnerID=40&md5=be26b5a33850da80a9da3633f1b78b40

Brans JP, Vincke P (1985) Note—a preference ranking organisation method. Manag Sci 31(6):647–656. doi:10.1287/mnsc.31.6.647. http://dx.doi.org/10.1287/mnsc.31.6.647

Brockett P, Cooper W, Golden L, Xia X (1997) A case study in applying neural networks to predicting insolvency for property and casualty insurers. J Oper Res Soc 48(12):1153–1162. doi:10.1057/palgrave.jors.2600461. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0031372229&partnerID=40&md5=ebd6a873b0d1549a6577f83e7d4f35be

Charnes A, Cooper W, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444. doi:10.1016/0377-2217(78)90138-8. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0018032112&partnerID=40&md5=a265c5a891d284d25f99f84c287298a1

Claveria O, Torra S (2014) Forecasting tourism demand to catalonia: neural networks vs. time series models. Econ Model 36:220–228

Corrente S, Greco S, Słowiński R (2013) Multiple criteria hierarchy process with ELECTRE and PROMETHEE. Omega (United Kingdom) 41(5):820–846. doi:10.1016/j.omega.2012.10.009. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84870996827&partnerID=40&md5=107c2904070cc7f319b2d7f29123ec53

Cortez P, Embrechts MJ (2013) Using sensitivity analysis and visualization techniques to open black box data mining models. Inf Sci 225:1–17

Creazza A, Restelli U, Porazzi E, Garagiola E, Croce D, Arpesella, M, Dallari F, Noè C (2016) Benchmarking of health technologies distribution models: an investigation of Lombardy's local health authorities. Benchmarking 23(4):817–842. doi:10.1108/BIJ-12-2013-0123. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84976349108&partnerID=40&md5=d83654195d28b220097480d3e365fda0

Daraio C, Simar L, Wilson P (2010) Testing whether two-stage estimation is meaningful in non-parametric models of production IAP stat technical report series (pp. 42): Institut de statistique. Universite Catholique de Louvain, Louvain-la-Neuve, Belgium

DiabetesUK (2016) What is diabetes? Report, Diabetes UK. https://www.diabetes.org.uk/Guide-to-diabetes/What-is-diabetes/. Accessed 15 July 2016

Efe B, Efe A (2016) An application of value analysis for lean healthcare management in an emergency department. Int J Comput Intell Syst 9(4):689–697. doi:10.1080/18756891.2016.1204117. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84980027712&partnerID=40&md5=4a0aba607c5144467ceefa9a868df1a2

Ertugrul I, Karakasoglu N (2009) Performance evaluation of Turkish cement firms with fuzzy analytic hierarchy process and topsis methods. Expert Syst Appl 36(1):702–715. doi:10.1016/j.eswa.2007.10.014. https://www.scopus.com/inward/record.uri?eid=2-s2.0-53849139805&partnerID=40&md5=dab7efb29718d78cfe6ff9f41b31b4d0

Faraway JJ (2016) Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models, vol 124. CRC, Boca Raton

Feng L, Zhang J (2014) Application of artificial neural networks in tendency forecasting of economic growth. Econ Model 40:76–80

Gallagher J (2015) Diabetes cases soar by 60% in past decade. BBC. http://www.bbc.co.uk/news/health-33932930. Accessed 29 Sep 2016

Gok MS, Altındağ E (2015) Analysis of the cost and efficiency relationship: experience in the Turkish pay for performance system. Eur J Health Econ 16(5):459–469. doi:10.1007/s10198-014-0584-6. http://dx.doi.org/10.1007/s10198-014-0584-6

Gomez A, Carnero M (2011) Selection of a computerised maintenance management system: a case study in a regional health service. Prod Plan Control 22(4):426–436. doi:10.1080/09537287.2010.500455. https://www.scopus.com/inward/record.uri?eid=2-s2.0-79957792317&partnerID=40&md5=57b23cdc9273ea7e31b84021c9b22800

Grigoroudis E, Zopounidis C (2012) Developing an employee evaluation management system: the case of a healthcare organization. Oper Res 12(1):83–106. doi:10.1007/s12351-011-0103-9. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84858291177&partnerID=40&md5=36c4cd3964255e42aa83e9a21ad7d569

Hadji B, Meyer R, Melikeche S, Escalon S, Degoulet P (2014) Assessing the relationships between hospital resources and activities: a systematic review. J Med Syst 38(10). doi:10.1007/s10916-014-0127-9. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84919916436&partnerID=40&md5=e296f95e5d1de968f98acc5213f21e1b

Hajek P, Michalak K (2013) Feature selection in corporate credit rating prediction. Knowl Based Syst 51:72–84. doi:10.1016/j.knosys.2013.07.008. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84883050191&partnerID=40&md5=45ef8680275354268ec902bc15c2495b

Hatami-Marbini A, Tavana M (2011) An extension of the electre I method for group decision-making under a fuzzy environment. Omega 39(4):373–386. doi:10.1016/j.omega.2010.09.001. https://www.scopus.com/inward/record.uri?eid=2-s2.0-78649901689&partnerID=40&md5=d62bc87bbe38ef08dbb0c009aff9adce

Hollingsworth B (2003) Non-parametric and parametric applications measuring efficiency in health care. Health Care Manag Sci 6(4):203–218. doi:10.1023/A:1026255523228. https://www.scopus.com/inward/record.uri?eid=2-s2.0-2942569893&partnerID=40&md5=029c925fb9190211b2ba45c402725d9c

Hwang CL, Yoon K (2012) Multiple attribute decision making: methods and applications a state-of-the-art survey, vol 186. Springer Science & Business Media, Berlin

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning, vol 6. Springer, Berlin

Kahraman H, Sagiroglu S, Colak I (2013) The development of intuitive knowledge classifier and the modeling of domain dependent data. Knowl Based Syst 37:283–295. doi:10.1016/j.knosys.2012.08.009. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84870062995&partnerID=40&md5=314c6d71104c0d59fb10c51a1735eacf

Kuhn M, Johnson K (2013) Applied predictive modeling. Springer, Berlin

Lai YJ, Liu TY, Hwang CL (1994) TOPSIS for MODM. Eur J Oper Res 76(3):486–500. doi:http://dx.doi.org/10.1016/0377-2217(94)90282-8. http://www.sciencedirect.com/science/article/pii/0377221794902828

Ledolter J (2013) Data mining and business analytics with R. Wiley, New York

McCullough JS, Crespin DJ, Abraham JM, Christianson JB, Finch M (2015) Public reporting and the evolution of diabetes quality. Int J Health Econ Manag 15(1):127–138

Mitropoulos P, Mitropoulos I, Sissouras A (2013) Managing for efficiency in health care: the case of greek public hospitals. Eur J Health Econ 14(6):929–938. doi:10.1007/s10198-012-0437-0. http://dx.doi.org/10.1007/s10198-012-0437-0

Moyo B (2012) Infrastructure quality and manufacturing exports in Africa: a firm-level analysis. S Afr J Econ 80(3):367–386. doi:10.1111/j.1813-6982.2011.01315.x. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84866232902&partnerID=40&md5=f4b55f5130aad5b3c45654bd0b388c43

Nguyen C, Pham N, Tran D, Lee A, Binns C (2016) Lifestyle and diet in relation to risk of type 2 diabetes in Vietnam: a hospital-based case—control study. SpringerPlus 5(1). doi:10.1186/s40064-016-2313-3. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84975807264&partnerID=40&md5=6a8f7f48046e511bf0376cb6199a7595

NHS (2013) Guide to the healthcare system in England: including the statement of NHS accountability. Report, NHS. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/194002/9421-2900878-TSO-NHS_Guide_to_Healthcare_WEB.PDF. Accessed 27 Aug 2016

Nuti S, Grillo Ruggieri T, Podetti S (2016) Do university hospitals perform better than general hospitals? a comparative analysis among italian regions. BMJ Open 6(8). doi:10.1136/bmjopen-2016-011426. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84982273230&partnerID=40&md5=270046c64a96300dd2e790b2cf6d5c6c

Opricovic S, Tzeng GH (2007) Extended vikor method in comparison with outranking methods. Eur J Oper Res 178(2):514–529. doi:10.1016/j.ejor.2006.01.020. https://www.scopus.com/inward/record.uri?eid=2-s2.0-33751011060&partnerID=40&md5=cb2c524791a035723e2b86d9afcc154f

Pelone F, Kringos D, Valerio L, Romaniello A, Lazzari A, Ricciardi, W., de Belvis A (2012) The measurement of relative efficiency of general practice and the implications for policy makers. Health Policy 107(2–3):258–268. doi:10.1016/j.healthpol.2012.05.005. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84865970504&partnerID=40&md5=a2b3508d3e870dc056d56c4c2f3d5b46

Pelone F, Kringos D, Spreeuwenberg P, De Belvis A, Groenewegen P (2013) How to achieve optimal organization of primary care service delivery at system level: lessons from europe. Int J Qual Health Care 25(4):381–393. doi:10.1093/intqhc/mzt020. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84876557289&partnerID=40&md5=2975e12e29ee6d64d35b39989d68c2d7

Roberts M (2015) Poor diabetes care in England 'costing lives'. Report. http://www.bbc.co.uk/news/health-30792848. Accessed 29 Aug 2016

Saaty T (1980) The analytical hierarchy process: planning, setting priorities, resource allocation. McGraw-Hill, New York

Seçme N, Bayrakdaroğlu A, Kahraman C (2009) Fuzzy performance evaluation in turkish banking sector using analytic hierarchy process and topsis. Expert Syst Appl 36(9):11699–11709. doi:10.1016/j.eswa.2009.03.013. https://www.scopus.com/inward/record.uri?eid=2-s2.0-67349187398&partnerID=40&md5=dbe1da9dd047eae71e408a5d599cc4b3

Siskos E, Askounis D, Psarras J (2014) Multicriteria decision support for global e-government evaluation. Omega 46:51–63. doi:10.1016/j.omega.2014.02.001. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84897808140&partnerID=40&md5=191fde2744df945e65afa08cf2544351

Sun J, Li H, Huang QH, He KY (2014) Predicting financial distress and corporate failure: a review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. Knowl Based Syst 57:41–56. doi:10.1016/j.knosys.2013.12.006. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84894901583&partnerID=40&md5=08855fd7c8b49859bdadeb41eda59553

Torgo L (2011) Data mining with R: learning with case studies. Chapman & Hall/CRC, Boca Raton, FL

Tsai WH, Chou WC, Hsu W (2009) The sustainability balanced scorecard as a framework for selecting socially responsible investment: an effective MCDM model. J Oper Res Soc 60(10):1396–1410. doi:10.1057/jors.2008.91. https://www.scopus.com/inward/record.uri?eid=2-s2.0-69649099483&partnerID=40&md5=cc42646f22d394aff388daa9943309da

Wang YM, Luo Y (2006) DEA efficiency assessment using ideal and anti-ideal decision making units. Appl. Math. Comput. 173(2):902–915. doi:10.1016/j.amc.2005.04.023. https://www.scopus.com/inward/record.uri?eid=2-s2.0-32644434592&partnerID=40&md5=87de38956a59ba955b067079d92af0fc

Wang B, Nistor I, Murty T, Wei YM (2014) Efficiency assessment of hydroelectric power plants in Canada: a multi criteria decision making approach. Energy Econ 46:112–121. doi:10.1016/j.eneco.2014.09.001. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84907749983&partnerID=40&md5=5e7432e2c9624c454a33b5a8f52d4185

Wang M, Fang H, Bishwajit G, Xiang Y, Fu H, Feng Z (2015) Evaluation of rural primary health care in Western China: a cross-sectional study. Int J Environ Res Public Health 12(11):13843–13860 (2015). doi:10.3390/ijerph121113843. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84946088473&partnerID=40&md5=55c380012d7d62fcf040a7f355c802e0

Wanke P, Azad M, Barros C (2016) Predicting efficiency in Malaysian islamic banks: a two-stage topsis and neural networks approach. Res. Int. Bus. Financ. 36:485–498. doi:10.1016/j.ribaf.2015.10.002. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84958580799&partnerID=40&md5=01c64a6104a8d22d182d528ce539f2c2

WHO (2016) Global report on diabetes. Report. http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf?ua=1. Accessed 26 Aug 2016

Wu D (2006) A note on dea efficiency assessment using ideal point: an improvement of Wang and Luo's model. Appl Math Comput 183(2):819–830. doi:10.1016/j.amc.2006.06.030. https://www.scopus.com/inward/record.uri?eid=2-s2.0-33845888809&partnerID=40&md5=96edec02de4be0144554d7ee2e94be9c

Yang CC (2016) Measuring health indicators and allocating health resources: a dea-based approach. Health Care Manag Sci 1–14. doi:10.1007/s10729-016-9358-2. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84957576629&partnerID=40&md5=0ba2fc595e3b9a0753c29f6817bec5bd

Yeh CC, Lin F, Hsu CY (2012) A hybrid KMV model, random forests and rough set theory approach for credit rating. Knowl Based Syst 33:166–172. doi:10.1016/j.knosys.2012.04.004. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84861589689&partnerID=40&md5=7c5ee7cb17921a0bdcdfa7775daca9d0

Chapter 15
A Multiobjective Solution Method for Radiation Treatment Planning

Gokhan Kirlik, Serpil Sayın, and Hao Howard Zhang

15.1 Introduction

Radiation therapy is to use various forms of radiation to shrink tumors and kill cancer cells for treatment of cancer (Mundt and Roeske 2003). One of the types of radiation therapy is called external beam radiation therapy (D'Amico et al. 1998), which uses computer-controlled devices to deliver high-energy radiation to the tumor from outside of the patient's body. The goal is to deliver right amount of radiation dose to the tumor while sparing the normal tissue and critical organs. Thus, the radiation treatment planning (RTP) requires to consider several conflicting objectives in nature. The resulting mathematical problem is a multiobjective optimization problem (MOP). Solution to an MOP consists of an efficient solution that has the property that no improvement on any objective is possible without deteriorating another one. Outcomes of such solutions are said to be nondominated. Several solution methods have been proposed to solve MOP (Ehrgott 2005). Most of these algorithms use a scalarization method which formulates a single-objective optimization problem in a way that the optimal solution to the single-objective optimization problem is an efficient solution to the MOP (Ehrgott 2006). Scalarization methods have been widely used in RTP to deal with multiple objectives (Craft 2013).

G. Kirlik (✉) • H.H. Zhang
Department of Radiation Oncology, University of Maryland School of Medicine, Baltimore, MD, USA
e-mail: gokhankirlik@umm.edu

S. Sayın
College of Administrative Sciences and Economics, Koç University, 34450 Sarıyer, Istanbul, Turkey

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_15

Most commonly used approach to obtain a treatment plan for multiobjective RTP is weighted sum method (Cotrutz et al. 2001; Lee et al. 2003; Breedveld et al. 2009) which is one of the scalarization methods (Ehrgott 2006). In this method, the MOP is transformed into a single objective problem by combining objective functions with positive weights (Zadeh 1963). However, for non-convex problems, some efficient solutions cannot be attained by using weighted sum scalarization (Ehrgott 2005). Here, the main problem is that the weight factors have ambiguous clinical meaning and that their relationship to the solution is not known in advance (Ehrgott et al. 2010). Therefore, the planner is required to often repeat the optimization with different factors to obtain some satisfactory solution. In order to overcome time consuming iterative processes in between the clinician and the treatment planning software, automated methods (Ahmed et al. 2010; Dias et al. 2016) and multiobjective solution methods (Craft et al. 2006; Thieke et al. 2007) were proposed. While the automated methods change the parameters iteratively without any user interaction, multiobjective solution methods create a database of plans by changing the weights systematically.

Thieke et al. (2007) develop an interactive algorithm for radiation therapy planning which is called multicriteria interactive radiotherapy assistant. This method computes $2^p - 1$ plans by solving lexicographic optimization problems for all possible combinations of objectives where p represents the number of objective functions. After computing the database, it is interactively explored by the planner until the satisfactory plan is obtained.

Craft et al. (2006) present an algorithm to create an approximation of the nondominated set for convex multiobjective RTP problems where nondominated set is the mapping of the efficient set in the outcome space. First, the method creates p plans by minimizing each objective function separately. Convex combination of these points forms the initial approximation of the nondominated set. Then, algorithm determines a weight vector and solves the weighted sum formulation to obtain a new treatment plan. The algorithm terminates when the stopping threshold for the approximation error is satisfied. The computational improvements of the algorithm is presented in Rennen et al. (2011). Bokrantz and Forsgren (2013) present an algorithm to approximate the convex nondominated sets and apply the algorithm to radiation therapy planning. Recently, Engberg et al. (2016) introduced a new problem formulation to improve the solution quality of the approximation. After creating the approximation, the final treatment plan is obtained by navigating over the approximation of the nondominated set (Monz 2006; Monz et al. 2008; Craft and Monz 2010). This is done in interaction with the decision maker (planner, clinician) to obtain a plan based on her/his preferences by navigating over the approximation (Hong et al. 2008). Since the plan is acquired by using the approximation set, it may not be a nondominated solution.

In this study, we use achievement scalarization to obtain deliverable plans considering several objectives for RTP (Wierzbicki 1980). Unlike weighted sum method, achievement scalarization is able to attain nondominated solutions from not only convex regions but also nonconvex regions of the nondominated set (Wierzbicki 1986, 1999). Here, our contribution lies in the way we define the

parameters of achievement scalarization subproblem. We first ask the clinicians to define a set of bounds on the objective functions. When we consider each structure (tumor volume, organs-at-risk and normal tissue) in the treatment volume as an objective function, dose bounds for that particular structure have clinical meaning and they are usually well-established through clinical protocols. Therefore they are relatively easy to specify. We then adapt the parameters of the scalarization to obtain a solution that is likely to satisfy the given bounds. We test our method with 10 locally advanced head-and-neck cancer cases, and compare the results with the plans that are created in our clinic and an algorithm from the literature (Bokrantz and Forsgren 2013). We observe that our achievement scalarization based method delivers significant improvements on the tested cases. In the following section, we present the background. In Sect. 15.3, the multiobjective programming formulation of the problem is presented. In Sect. 15.4, the proposed method is tested on clinical data. Finally, conclusions are presented in Sect. 15.5.

15.2 Background

In MOP, p objective functions $f_j(x) : \mathbb{R}^n \to \mathbb{R}$ for $j = 1, \ldots, p$ have to be minimized. The feasible set is denoted as $\mathscr{X} \subseteq \mathbb{R}^n$. Each feasible solution $x \in \mathscr{X}$ is mapped into its corresponding objective vector $y = f(x)$ and $\mathscr{Y} = \{y \in \mathbb{R}^p : y = f(x)$ for some $x \in \mathscr{X}\}$ is referred to as the set of feasible outcomes in the objective space. In mathematical terms, MOP is defined as:

$$\begin{aligned} \text{(MOP)} \quad & \min f(x) = [f_1(x), \ldots, f_p(x)]^T \\ & \text{s.t.} \quad x \in \mathscr{X} \end{aligned}$$

Due to conflicting objectives, MOP problem is expected to have more than one solution. These solutions are called efficient solutions.

Definition 1 A solution $x^* \in \mathscr{X}$ is called an **efficient solution** if there exists no feasible solution $x \in \mathscr{X}$ such that $f_j(x) \leq f_j(x^*)$ for all $j = \{1, \ldots, p\}$ and there exists $\hat{\jmath} \in \{1, \ldots, p\}$ such that $f_{\hat{\jmath}}(x) < f_{\hat{\jmath}}(x^*)$. For an efficient solution $x^*, f(x^*) \in \mathbb{R}^p$ is referred to as a **nondominated solution** in the outcome space.

In this study, we use augmented achievement scalarization (Wierzbicki 1980) to obtain an efficient solution for the MOP. Given positive weight $w \in \mathbb{R}^p$, reference point $r \in \mathbb{R}^p$ and a sufficiently small positive constant $\rho \in \mathbb{R}$, the augmented achievement scalarization is defined as follows,

$$\boldsymbol{P(w,r)} \quad \min_{x \in \mathscr{X}} \left\{ \max_{j=1,\ldots,p} w_j(f_j(x) - r_j) + \rho \sum_{j=1}^{p} w_j(f_j(x) - r_j) \right\}. \tag{15.1}$$

Theorem 1 shows that for any $r \in R^p$ and any positive $w \in \mathbb{R}^p$, the optimal solution of $P(w, r)$ is an efficient solution without convexity assumption.

Theorem 1 *Given a reference point $r \in \mathbb{R}^p$, a positive weight vector $w \in \mathbb{R}^p$ and sufficiently small positive scalar $\rho \in \mathbb{R}$, an optimal solution to $P(w, r)$ is efficient.*

Proof Let $w, r \in \mathbb{R}^p$ w positive, be given. Let x^* be an optimal solution to $P(w, r)$. Suppose x^* is not efficient. Then there exists a solution $x' \in \mathscr{X}$ such that $f_j(x') \leq f_j(x^*)$ for all $j \in \{1, \ldots, p\}$ and $f_{\hat{j}}(x') < f_{\hat{j}}(x^*)$ for some $\hat{j} \in \{1, \ldots, p\}$. It follows that $w_j(f_j(x') - r_j) \leq w_j(f_j(x^*) - r_j)$ for all $j \in \{1, \ldots, p\}$ and $\sum_{j=1}^{p} w_j(f_j(x') - r_j) < \sum_{j=1}^{p} w_j(f_j(x^*) - r_j)$. This implies that, $\max_{j=1,\ldots,p} w_j(f_j(x') - r_j) + \rho \sum_{j=1}^{p} w_j(f_j(x') - r_j) < \max_{j=1,\ldots,p} w_j(f_j(x^*) - r_j) + \rho \sum_{j=1}^{p} w_j(f_j(x^*) - r_j)$. This contradicts optimality of x^* to $P(w, r)$. Therefore x^* must be an efficient solution. □

In this study, we aim to target a nondominated solution in between the given lower and upper bounds on the objective functions. In general, finding a nondominated solution in between given bounds is a bilevel optimization problem which is an $\mathscr{NP}$-hard problem (Kirlik 2014). Therefore, instead of using a global optimization problem, we use achievement scalarization to target a nondominated solution in between given bounds. Let the lower bound and upper bound vectors on the objective functions be denoted as $l \in \mathbb{R}^p$ and $u \in \mathbb{R}^p$, respectively. Since we do not know the nondominated solutions apriori, it is a challenging problem to determine the parameters of the achievement scalarization formulation. Here, we determine the parameters of the achievement scalarization by using the target lower and upper bounds defined by the decision maker. As a reference point for the achievement scalarization method, we use upper bound vector, i.e. $r_j = u_j$ for all $j \in \{1, \ldots, p\}$. The weight vector of the achievement scalarization is determined as $w_j = 1/(u_j - l_j)$ for all $j \in \{1, \ldots, p\}$. These parameters, i.e. $w_j = 1/(u_j - l_j)$ and $r_j = u_j$ for all $j \in \{1, \ldots, p\}$, do not guarantee to find an efficient solution that maps into these bounds. Still, the problem formulation has an incentive to find one and in our tests we observe that the delivered solutions meet the expectations.

15.3 Problem Formulation

In Intensity Modulated Radiation Therapy (IMRT), we have a set of beam angles which is denoted by θ. From each beam angle, the linear accelerator delivers radiation which passes through a multileaf collimator (MLC), a device that consists of a number of pairs of metal leaves moving along channels into the path of radiation to block out areas of the beam. By moving the leaves into different positions, intensity modulation and beam shaping can be achieved. An MLC unit is shown on Fig. 15.1. This figure also shows how MLC unit shapes the beam. The set of apertures (different shapes of MLC) that has been used from beam angle $i \in \theta$ is denoted by K_i.

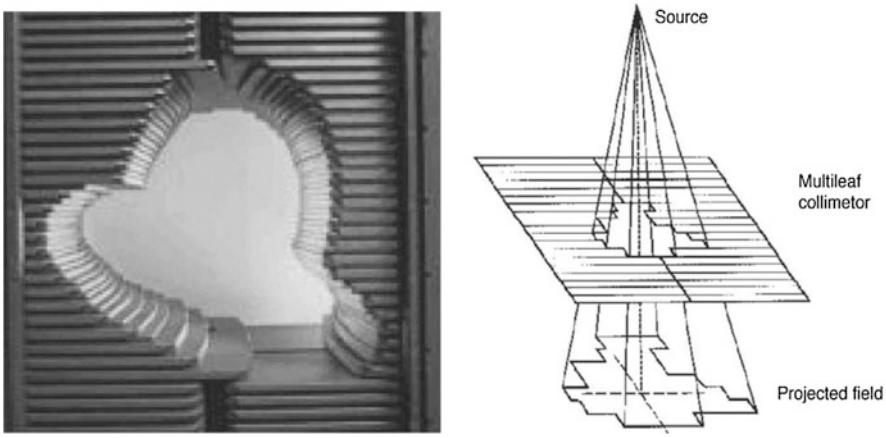

Fig. 15.1 An MLC unit and shaping the beam by using the MLC

In this study, we assume that the set of apertures from each angle is given. Hence, the decision variables of the problem are the weights associated with the dose delivered to aperture-k where $k \in K_i$ and $i \in \theta$. This problem is referred to as segment weight optimization problem. We present a multiobjective linear programming formulation of segment weight optimization problem for IMRT.

Before we introduce our formulation, we describe the essential aspects of segment weight optimization problem. The treatment region is divided into sub-regions. The set of all structures in the treatment volume is denoted by S. $T \subseteq S$ represents the set of target structures, $O \subseteq S$ denotes the set of organs-at-risk (OAR), $C \subseteq S$ denotes the set of critical structures. Each of the structures s is decomposed into a finite number of cubes which are called voxels. The set of voxels of structure s is denoted by V_s where $s \in S$. $D_{ik\ell}$ denotes the dose received by voxel $\ell \in V_s$ from aperture $k \in K_i$ of beam angle $i \in \theta$. The matrix D is called as dose deposition matrix. Given the properties of the patient's anatomy, source and angle, the values of the dose deposition matrix are calculated by mathematical models of the physical behavior of radiation as it travels through the body (Ehrgott et al. 2010), which constitutes an important parameter of our formulation. With the given notation and parameters, the total amount of dose received by voxel-ℓ, denoted z_ℓ, is computed as follows,

$$z_\ell = \sum_{i \in \theta} \sum_{k \in K_i} b_{ik} D_{ik\ell} \quad \ell \in V_s,\ s \in S. \tag{15.2}$$

In (15.2), b_{ik} is the decision variable which determines the weight of the beam from angle $i \in \theta$ delivered to aperture $k \in K_i$.

In our multiobjective optimization formulation, we have three sets of objective functions originating from target structures, OARs and critical structures. The first set of objectives is for regulating the dose received by targeted tumors. Each tumor

structure has its own prescription dose where prescription dose to $s \in T$ is denoted as PD_s. For the target structures, the aim is to deliver a uniform prescription dose as much as possible. Hence the first set objective functions consist of maximum variation from the prescription dose. This can be expressed as

$$f_s(z) = \max_{\ell \in V_s} |z_\ell - PD_s| \quad s \in T. \tag{15.3}$$

For the OARs, the general aim is to minimize the mean dose they receive. In our second set of objective functions, the mean dose for each OAR is formulated as follows,

$$f_s(z) = \frac{1}{|V_s|} \sum_{\ell \in V_s} z_\ell \quad s \in O. \tag{15.4}$$

In (15.4), $|V_s|$ denotes the number of voxels in structure s.

The third set of objective functions is for the critical organs such as the brain and the spinal cord. Every structure in this set, i.e. $s \in C$, is a serial organ, so the aim is to minimize the maximum dose delivered to these organs which is expressed as follows,

$$f_s(z) = \max_{\ell \in V_s} z_\ell \quad s \in C. \tag{15.5}$$

Thus, our formulation becomes

$$\min \quad f_s(z) \qquad\qquad s \in S \tag{15.6}$$

$$s.t. \quad \zeta_s - \frac{1}{(1-\alpha_s)|V_s|} \sum_{\ell \in V_s} \max(\zeta_s - z_\ell, 0) \geq PD_s \qquad s \in T \tag{15.7}$$

$$z_\ell = \sum_{i \in \theta} \sum_{k \in K_i} b_{ik} D_{ik\ell} \qquad \ell \in V_s,\ s \in S \tag{15.2}$$

$$b_{ik} \geq 0 \qquad i \in \theta,\ k \in K_i \tag{15.8}$$

$$\zeta_s \text{ is free} \qquad s \in T \tag{15.9}$$

The only constraint in our formulation seeks to guarantee the dose delivered to the tumor structures. At least a certain fraction, α_s for $s \in T$, of any planning target volume (PTV) should receive the prescription dose. We use conditional value-at-risk (CVaR) constraints to guarantee that at least α_s percentage of any $s \in T$ receives PD_s (Romeijn et al. 2003).

15.4 Results

We tested our approach on 10 locally advanced head-and-neck cancer cases. Figure 15.2 shows two of the cases as illustration. Locally advanced head-and-neck cancer is one of the most challenging disease cites for external beam radiation therapy. The treatment targets involve primary tumors and nodal volumes. The treatment plans were created by using Pinnacle[3] (Philips Healthcare 2015). Nine beam angles were used in all treatment plans, $\theta = \{0, 40, 80, 120, 160, 200, 240, 280, 320\}$.

All cases include three tumor volumes in target, which are primary tumor with positive nodes, high-risk nodal volume and low-risk nodal volume, $|T| = 3$. The OARs included the left and right parotid glands and oral cavity, $|O| = 3$. Finally, critical organs are spinal cord and brain stem, $|C| = 2$. Total number of voxels in the treatment volume, $\sum_{s \in S} |V_s|$, and total number of apertures, $\sum_{i \in \theta} |K_i|$ where $|K_i|$ is the number of apertures from beam angle $i \in \theta$, of each case are given in Table 15.1. Note that total number of voxels and total number of apertures reflect the size of the optimization problem.

In the optimization process, we used previously published constraints (Zhang et al. 2013). In addition to the specifications addressed in constraints, the proposed method requires lower and upper bounds on the objective functions in order to formulate the achievement scalarization function. We defined these bounds by using dose and dose volume goals, specified in Gy[1] for locally advanced head and neck cases which are given in Table 15.2. The prescription doses (PD_s for $s \in T$) are

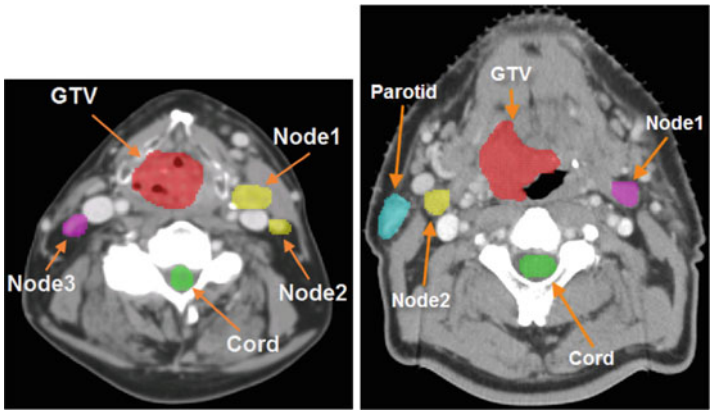

Fig. 15.2 Axial slices of two locally advanced head-and-neck cases with gross tumor volume, nodal volumes and OARs

[1]The gray (Gy) is a derived unit of ionizing radiation dose in the International System of Units (SI). One gray is the absorption of one joule of energy, in the form of ionizing radiation, per kilogram of matter (Taylor 1995).

Table 15.1 Problem size of 10 head-and-neck cases

	# of voxels	# of apertures
Case-1	26,668	329
Case-2	12,552	313
Case-3	15,291	269
Case-4	9948	171
Case-5	24,757	263
Case-6	19,415	306
Case-7	11,641	220
Case-8	17,383	284
Case-9	13,624	260
Case-10	14,620	277

Table 15.2 Summary of planning goals for locally advanced head and neck cases

Target volume	
% volume of primary PTV receiving $\geq$ 70 Gy	95
% volume of primary PTV receiving $\geq$ 77 Gy	5
% volume of high-risk PTV receiving $\geq$ 59.4 Gy	95
% volume of high-risk PTV receiving $\geq$ 65 Gy	5
% volume of low-risk PTV receiving $\geq$ 54 Gy	95
% volume of low-risk PTV receiving $\geq$ 59.4 Gy	5
Organs-at-risk	
Left parotid mean dose	26 Gy
Right parotid mean dose	26 Gy
Oral cavity mean dose	35–40 Gy
Critical organs	
Maximum dose for spinal cord	45 Gy
Maximum dose for brain stem	54 Gy

Table 15.3 Dose bounds on each structure in the treatment volume

Maximum variation from the prescription dose:	$[0, 0.1\, PD_s]$ Gy for all $s \in T$
Left parotid mean dose:	[20, 26] Gy
Right parotid mean dose:	[20, 26] Gy
Oral cavity mean dose:	[35, 40] Gy
Maximum dose for spinal cord:	[40, 45] Gy
Maximum dose for brain stem:	[50, 54] Gy

70 Gy to the primary tumor, 59.4 Gy to the high-risk nodal volume and 54 Gy to the low-risk nodal volume. The lower and upper bounds on the objective functions are given in Table 15.3.

We also compare our method with Bokrantz and Forsgren's algorithm (BF method) (Bokrantz and Forsgren 2013) that utilizes database of plans to obtain the final treatment plan. This method is also a multiobjective optimization solver of a well-known treatment planning system (RaySearch Laboratories 2015). First,

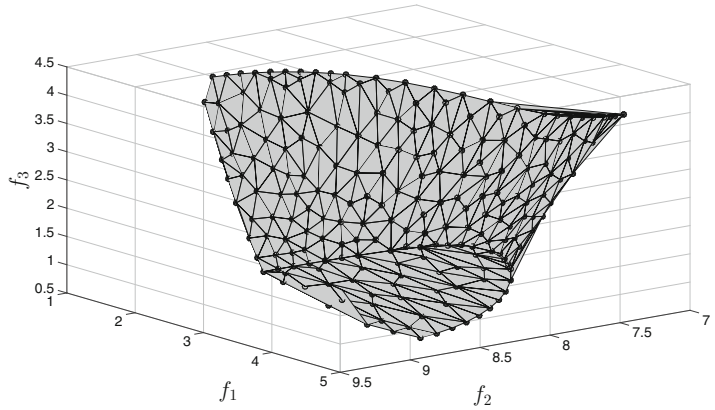

Fig. 15.3 An approximation of the nondominated set of a three-objective convex optimization problem which is obtained by using BF method

the method creates a set of solutions for the MOP based on decision maker's error tolerance. By using these solutions, method generates an approximation for the nondominated set. In Fig. 15.3, we show the set of discrete solutions and approximation set of a three-objective convex optimization problem which is obtained by using BF method. Here, the convex optimization problem is defined as $\min\{[x_1, \ldots, x_p] : \sum_{j\in\{1,\ldots,p\}\setminus\{i\}}(x_j - a_j)^2 - x_i \leq 0,\ i = 1, \ldots, p\}$ where $p = 3$. In Fig. 15.3, dots are the nondominated solutions which are obtained by using weighted sum formulation, and the union of faces form the approximation of the nondominated set. After obtaining the approximation set, the method navigates over the approximation of the nondominated set to obtain the final treatment plan.

Both the BF method and the proposed method are implemented in C++ and linear programming subproblems are solved by using IBM CPLEX 12.6 (IBM 2014). Since the subproblems are large (see Table 15.1) and sparse, we use the Barrier method instead of the primal or dual simplex methods. For the BF method, we use the same multiobjective linear programming formulation with the proposed method, i.e. the formulation given in Sect. 15.3, to compare the methodological difference between two methods. Treatment volume includes eight structures. Hence, the multiobjective formulation contains eight objective functions, i.e. $p = 8$. For the BF method, we need to define the number of discrete plans that is used to create the approximation. There is a trade-off between the number of plans used in the approximation and the solution time (Craft and Bortfeld 2008). The method requires at least $p + 1$ number of plans. For the BF method, we created $3 \times p$ plans to obtain the approximation. BF method also requires an algorithm for the navigation over the approximation of the nondominated set. We use the method presented by Craft and Monz (2010) which is based on Pascoletti and Serafini scalarization (Pascoletti and Serafini 1984).

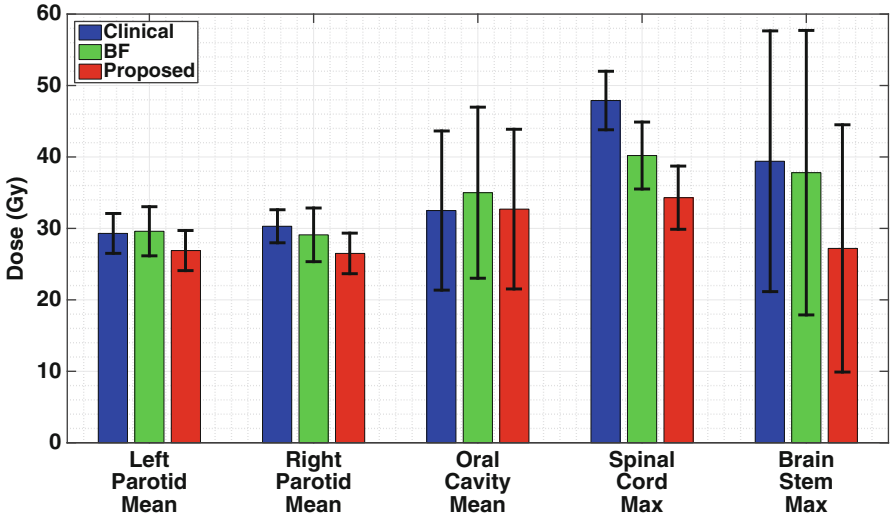

Fig. 15.4 Achieved doses (mean ± standard deviation in Gy) at constraint setting for OARs by clinical, BF and proposed method plans

Table 15.4 Comparing the average dose differences achieved by three methods: clinical treatment plan, BF method and our proposed method

	Left parotid mean	Right parotid mean	Oral cavity mean	Spinal cord max	Brain stem max
Clinical-BF	1.18%	−3.63%	7.06%	−15.04%	−8.53%
p-value	0.64	0.18	0.97	< 0.01	0.36
Clinical-proposed	−8.03%	−12.48%	−0.03%	−27.83%	−35.90%
p-value	< 0.01	< 0.01	0.55	< 0.01	< 0.01
BF-proposed	−8.92%	−7.89%	−6.95%	−13.12%	−24.29%
p-value	< 0.01	0.03	< 0.01	0.02	0.01

We compare the proposed algorithm with the clinical plans and the BF method. For final plan comparison, the clinical plans were normalized so that at least 95% of each target received the corresponding prescription dose. Note that, the plans created by using the BF method and proposed method do not require normalization, since the problem formulation includes dose volume constraint for each PTV. Since all plans provided sufficient coverage for all three targets, we may compare all plans in terms of OARs and critical organ toxicity. Figure 15.4 uses a bar chart to compare the mean doses for both parotid glands and oral cavity, and the maximum doses for the spinal cord and brain stem. The average of the ten cases and the standard deviation are displayed. Reduction in doses received by OARs was achieved with proposed method compared to both clinical plan and BF method. Table 15.4 shows the details. p-value of paired t-test in Table 15.4 shows the difference in achieved

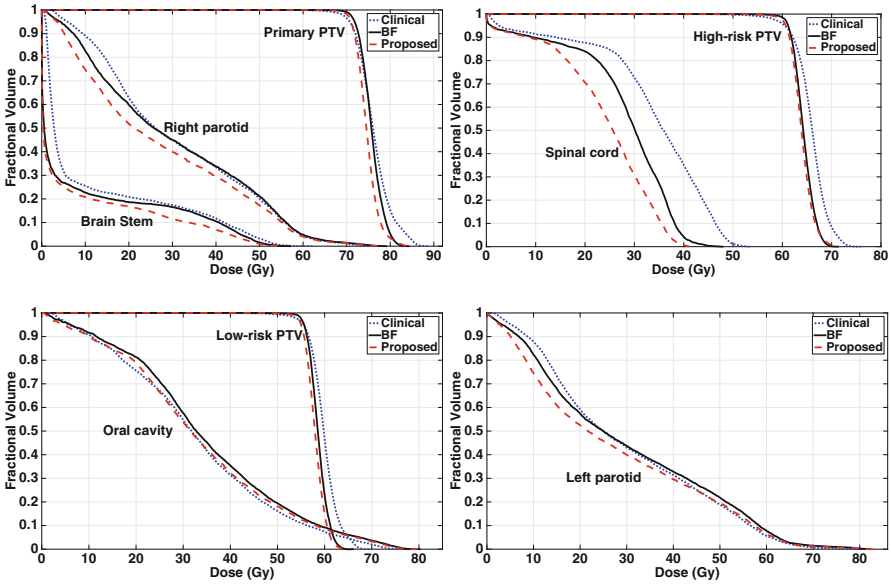

Fig. 15.5 Average DVHs for Clinical, BF method and proposed method plans

dose was significant (p-value < 0.01). The proposed method plans outperformed the clinical plans and the plans created by using BF method. Average reductions by the proposed method are 8.03%, 12.48%, 0.03%, 27.83% and 35.90% as compared to clinical IMRT plans and 8.92%, 7.89%, 6.95%, 13.12% and 24.29% as compared to BF method plans for the left parotid, right parotid, oral cavity, spinal cord and brain stem respectively.

For a better comparison, we use dose-volume histogram (DVH) to compare three sets of plans where DVH is a histogram relating radiation dose to tissue volume in radiation therapy planning (Drzymala et al. 1991). The mean DVHs of the clinical plans, BF method plans and proposed plans are shown in Fig. 15.5. We can see that all plans provided sufficient coverage for all three targets. Additionally, the plans that are obtained by the proposed method provide more uniform dose to PTV compared to clinical plans and BF method plans. The improvement in the OAR sparing with the proposed plans compared to clinical plans and BF method can also be seen in Fig. 15.5.

Finally, we compare the CPU time of BF method and the proposed method. The CPU time comparison results are reported in Table 15.5. Since the BF method creates database of plans, its CPU time is significantly larger than the proposed method. Still, CPU time of the BF method can be decreased by sacrificing the quality of the approximation. As seen in Table 15.5, the CPU time of the proposed method varies in between 2 and 3 min. This means that our method is able create high quality plans without any interaction with the planner in a short time.

Table 15.5 CPU time comparison of the BF method and the proposed method in CPU seconds

	BF method	Proposed method
Case-1	2985.6	201.7
Case-2	40,993.6	83.4
Case-3	1911.1	118.1
Case-4	5484.7	44.0
Case-5	2392.5	149.6
Case-6	2461.1	126.4
Case-7	12,730.0	66.9
Case-8	1597.8	95.4
Case-9	20,836.4	74.5
Case-10	4939.0	105.3
Average	9633.2	106.5

15.5 Conclusions

We present a novel multiobjective solution approach for segment weight optimization for IMRT. The proposed method eliminates the interaction with the decision maker while creating the final treatment plan compared to previous methods. We tested our method with 10 locally advanced head-and-neck cancer cases. Since the method considers multiple objectives while creating the final treatment plan, the method was able to generate high quality plans compared clinical plans. The proposed method also outperformed a multiobjective optimization solver of a well-known treatment planning system. Proposed method was able to decrease the amount of dose delivered to OAR up to 35% and 24% compared to clinical plan and the multiobjective optimization solver. As a future work, instead of fixed set of apertures, we will generate deliverable apertures considering multiple objectives.

References

Ahmed S, Gozbasi O, Savelsbergh M, Crocker I, Fox T, Schreibmann E (2010) An automated intensity-modulated radiation therapy planning system. INFORMS J Comput 22(4):568–583

Bokrantz R, Forsgren A (2013) An algorithm for approximating convex pareto surfaces based on dual techniques. INFORMS J Comput 25(2):377–393

Breedveld S, Storchi PR, Heijmen BJ (2009) The equivalence of multi-criteria methods for radiotherapy plan optimization. Phys Med Biol 54(23):7199

Cotrutz C, Lahanas M, Kappas C, Baltas D (2001) A multiobjective gradient-based dose optimization algorithm for external beam conformal radiotherapy. Phys Med Biol 46(8):2161

Craft D (2013) Multi-criteria optimization methods in radiation therapy planning: a review of technologies and directions. arXiv preprint arXiv:13051546

Craft D, Bortfeld T (2008) How many plans are needed in an IMRT multi-objective plan database? Phys Med Biol 53(11):2785

Craft D, Monz M (2010) Simultaneous navigation of multiple Pareto surfaces, with an application to multicriteria IMRT planning with multiple beam angle configurations. Med Phys 37(2):736–741

Craft DL, Halabi TF, Shih HA, Bortfeld TR (2006) Approximating convex Pareto surfaces in multiobjective radiotherapy planning. Med Phys 33(9):3399–3407

D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, Tomaszewski JE, Renshaw AA, Kaplan I, Beard CJ et al (1998) Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. J Am Med Assoc 280(11):969–974

Dias J, Rocha H, Ventura T, Ferreira B, do Carmo Lopes M (2016) Automated fluence map optimization based on fuzzy inference systems. Med Phys 43(3):1083–1095

Drzymala R, Mohan R, Brewster L, Chu J, Goitein M, Harms W, Urie M (1991) Dose-volume histograms. Int J Radiat Oncol Biol Phys 21(1):71–78

Ehrgott M (2005) Multicriteria optimization. Springer, Berlin

Ehrgott M (2006) A discussion of scalarization techniques for multiple objective integer programming. Ann Oper Res 147(1):343–360

Ehrgott M, Güler Ç, Hamacher HW, Shao L (2010) Mathematical optimization in intensity modulated radiation therapy. Ann Oper Res 175(1):309–365

Engberg L, Eriksson K, Forsgren A, Hårdemark B (2016) Explicit and convex optimization of plan quality measures in intensity-modulated radiation therapy treatment planning. arXiv preprint arXiv:160701148

Hong TS, Craft DL, Carlsson F, Bortfeld TR (2008) Multicriteria optimization in intensity-modulated radiation therapy treatment planning for locally advanced cancer of the pancreatic head. Int J Radiat Oncol Biol Phys 72(4):1208–1214

IBM (2014) IBM ILOG Cplex Optimization Studio 12.6

Kirlik G (2014) Exact and representation methods for multiobjective optimization problems. PhD thesis, Koç University

Lee EK, Fox T, Crocker I (2003) Integer programming applied to intensity-modulated radiation therapy treatment planning. Ann Oper Res 119(1–4):165–181

Monz M (2006) Pareto Navigation–interactive multiobjective optimisation and its application in radiotherapy planning. PhD thesis, Department of Mathematics, Technical University of Kaiserslautern, Germany

Monz M, Küfer K, Bortfeld T, Thieke C (2008) Pareto navigation—algorithmic foundation of interactive multi-criteria IMRT planning. Phys Med Biol 53(4):985

Mundt AJ, Roeske JC (2003) Principles of radiation oncology. In: Oncologic therapies. Springer, Berlin pp 9–17

Pascoletti A, Serafini P (1984) Scalarizing vector optimization problems. J Optim Theory Appl 42(4):499–524

Philips Healthcare (2015) Pinnacle^3

RaySearch Laboratories (2015) Multicriteria optimization in RayStation. White paper

Rennen G, van Dam ER, Hertog Dd (2011) Enhancement of sandwich algorithms for approximating higher-dimensional convex pareto sets. INFORMS Journal on Computing 23(4):493–517

Romeijn HE, Ahuja RK, Dempsey JF, Kumar A, Li JG (2003) A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning. Phys Med Biol 48(21):3521

Taylor B (1995) Guide for the use of the International System of Units (SI): the metric system. DIANE Publishing, Collingdale

Thieke C, Küfer KH, Monz M, Scherrer A, Alonso F, Oelfke U, Huber PE, Debus J, Bortfeld T (2007) A new concept for interactive radiotherapy planning with multicriteria optimization: first clinical evaluation. Radiother Oncol 85(2):292–298

Wierzbicki AP (1980) The use of reference objectives in multiobjective optimization. In: Fandel G, Gal T (eds) Multiple criteria decision making theory and application. Lecture notes in economics and mathematical systems, vol 177. Springer, Berlin, Heidelberg, pp 468–486

Wierzbicki AP (1986) A methodological approach to comparing parametric characterizations of efficient solutions. In: Large-scale modelling and interactive decision analysis. Springer, Berlin, pp 27–45

Wierzbicki AP (1999) Reference point approaches. In: Multicriteria decision making. Springer, Berlin, pp 237–275

Zadeh L (1963) Optimality and non-scalar-valued performance criteria. IEEE Trans Autom Control 8(1):59–60

Zhang H, Betzel G, Yi B, D'Souza W (2013) Beam controlled arc therapy–a delivery concept for stationary targets. Phys Med Biol 58(20):7117

Part VII
Medical Issues

Chapter 16
OR Models for Emergency Medical Service (EMS) Management

S.M. Gholami-Zanjani, M.S. Pishvaee, and S. Ali Torabi

16.1 Introduction to Emergency Medical Services

Emergency medical services (EMS) is a comprehensive procedure for prompt patient care in different settings such as the emergency departments (ED), urgent care clinics, pre-hospital settings, and also disaster situations (Soriya and Colwell 2012). EMS typically consists of call centers, dispatchers, ambulances and paramedics. Accordingly, process of EMS follows some steps such as reception of calls, vehicle dispatching, on-site treatment and release of patient or transportation to health care facilities. This process is presented in Fig. 16.1. Therefore, emergency medical service plays a critical role in any emergency care system. The growing demand for more efficient EMS has sparked off efforts to evaluate and improve the quality of many EMS systems in recent years. The key objective of many successful EMS systems can be operationally defined by the effective and consistent provision of immediate medical care to seriously ill or injured patients, and the expeditious conveyance of patients to advanced resuscitation and lifesaving care (I.O.M 2006). Medical conditions of patients who require immediate care imply that delay in treatment could result in the worsening of patients' conditions (Lam et al. 2015).

All abovementioned issues indicate the importance of EMS area. Among several relevant issues in this context, ambulance planning problem could help to improve public health level and social health care systems. Generally, the problem of

S.M. Gholami-Zanjani • M.S. Pishvaee
School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran
e-mail: se_Gholami@ind.iust.ac.ir; pishvaee@iust.ac.ir

S.A. Torabi (✉)
School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran
e-mail: satorabi@ut.ac.ir

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_16

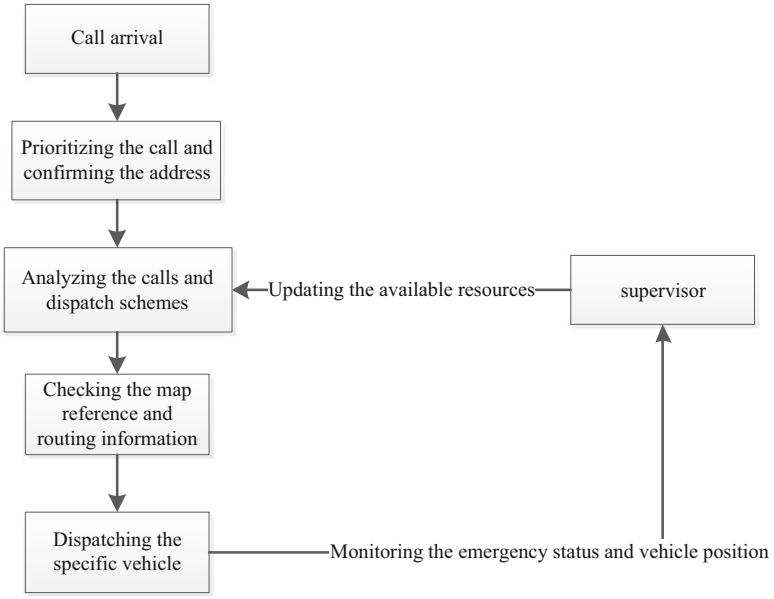

Fig. 16.1 Emergency response process (Adopted from Haghani and Yang 2007)

locating ambulances and ambulance bases can be divided into strategic, tactical and operational levels. At the strategic level, the locations of the ambulance bases are determined while considering coverage constraints (Berman et al. 2011; Karasakal and Karasakal 2004; Brotcorne et al. 2003; Schmid 2012; Schmid and Doerner 2010). In the next step, the required number of ambulances to fulfill the demand of each base is specified over a multi-period mid-term decision horizon (Brotcorne et al. 2003; Schmid 2012; Essen et al. 2013; Knight et al. 2012; Galvão et al. 2005). Finally, at the operational level, the allocation of ambulances to emergency demands and relocation of ambulances must be carried out in the real-time in an on-going fashion. Therefore, the main decisions to be made involve: the number of required resources, ambulance/helicopter fleet management, location of ambulances, and ambulance routing. The important point is that all aforementioned decisions are highly interrelated and their joint optimization in a sole model could prevent sub-optimality of obtained solutions (Coppi et al. 2013; Jagtenberg et al. 2015).

Dynamic and complex nature of emergency medical systems imposes a high degree of uncertainty in ambulance planning decisions and significantly influences the overall performance of the achieved decisions. Some examples of uncertainties in this context are randomness in the demand of emergency services, location of incidents and variability of travel times. Thus, the significance of accounting for uncertainty has prompted the researchers to address uncertain parameters in EMS models. Stochastic programming approach is often employed by researchers to cope with data uncertainty in the proposed models (Essen et al. 2013; Galvão et al. 2005).

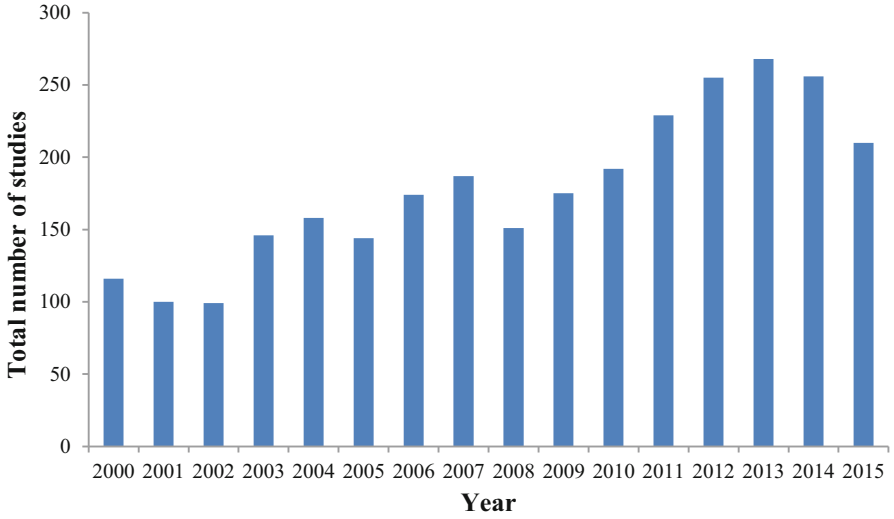

Fig. 16.2 General trend of healthcare problems

Having reviewed the academic literature of EMS, the general trend of conducted studies, in terms of number of papers, is extracted. In this regard, we have investigated the research titles indexed in Google Scholar and Scopus. These research subjects include "location", "emergency", "ambulance", "routing", "hospital" and "medical" with various combinations in the titles and subjects. All relevant studies published since 2000 have been reviewed. Being mindful of the fact that we may miss some indirect relevant papers, overlapping some researches and considering papers that could not be directly involved in the investigated framework, the resulted papers could fairly mirror the general trend. All together, we have ended up with numerous papers in this field since 2000. Figure 16.2 shows the number of articles with respect to relevant years. It is obvious that there is a growing interest in the subject. It seems that this trend in drawing attentions is likely to continue due to rise in the elderly population of the societies.

As it is mentioned earlier, the EMS problems could be classified into two main categories: (1) strategic and (2) tactical/operational level problems. Strategic decisions are mainly concerned with location of ambulance stations (i.e. those sites where ambulances are waiting to be dispatched to incident locations), medical care facilities and staff hiring while tactical/operational decisions focus on short and mid-term decisions such as crew scheduling, crew pairing, number of needed vehicles, ambulance standby site location, ambulance relocation and dispatching rules. Dispatching decisions are related to ambulance routing problems. Generally, the main three types of decisions in the emergency medical service include: (1) ambulance location, (2) ambulance relocation, and (3) ambulance dispatching, which significantly influence the cost efficiency and responsiveness (e.g. response

time) measures (Bruzzone and Signorile 1998). In this chapter, we focus on strategic and operational decisions as per their importance in this area.

The remainder of this chapter is continued in Sect. 2 by introducing and reviewing EMS strategic planning models. EMS operational planning models are described and investigated in Sect. 3. These sections include distinguished models from the related literature as well. Section 3.1 describes how different uncertainty programming approaches are applied to EMS models. An application of EMS models on a real case is illustrated in Sect. 5. Finally, Sect. 6 includes concluding remarks and some future research directions.

16.2 EMS Strategic Planning Models

In EMS systems, response time is considered as a critical issue that directly affects patients' health. Insufficient prompt services dramatically threaten people's lives. The public expects availability and accessibility of EMS facilities to provide sufficient services (Daskin and Dean 2004). However, because of limited allocated governmental budget and other resources, the quality and quantity of services could be vulnerable. By availability we mean adequate supply of services. The opportunity to obtain medical services and resources is defined as the access to healthcare facilities (Gulliford et al. 2002). In other words, accessibility of facilities refers to the capability of the patients to reach the health care facilities. In the case of prompt response services, it is vice versa; capability of health providers such as ambulances to reach patients is considered as the accessibility. In order to attain the desired service level, location of health care facilities such as ambulance bases should be efficiently determined. In addition, high cost resources are involved in maintaining EMS in terms of ambulances, qualified staff and so on (Farahani et al. 2012). Hence, effectively locating the EMS facilities becomes an impressive decision on the system configuration. There are three common service goals used in EMS which are as follows (Kerkkamp 2014).

- Minimizing the average response time to patients' needs
- Minimizing the maximum response time to patients' needs
- Maximizing the number of regions that are supportable (i.e. covered) within a predefined response time

These three goals are corresponding to three basic location models namely, the P-median, P-center and the covering models, respectively. Throughout this chapter our attention is restricted to discrete location models because they have been used considerably in health care location problems while they are practical to reasonable extent. Covering models are used abundantly in the related literature as they are suitable enough for emergency situations. There are several comprehensive reviews that provide deep insights and state-of-the-art in this field (Brotcorne et al. 2003; Farahani et al. 2012; Goldberg 2004; Li et al. 2011; Looije 2013; Van Buuren 2011).

The classical Location Set Covering Model (LSCM) and Maximal Covering Location Problem (MCLP) introduced by Toregas et al. (1971) and Church and ReVelle (1974) are two variants of covering models. These models assume that all the parameters are deterministic and a demand point is roughly covered or uncovered by each located facility. By covering we mean that a node or a demand point is covered by a facility if the distance or response time between the facility and the demand point is less or equal to a predefined value. The sets, inputs and variables used to define the LSCM are as follows.

Set	***Description***
I	Set of demand nodes
J	Set of potential facility sites
f_j	Fixed cost of locating a facility at site j
a_{ij}	Equals to 1 if demand point j can be covered by base (i.e. location) i, otherwise it equals to 0

Variable	***Description***
X_j	Equals to 1 if a facility is located at potential site j, otherwise 0

With this notation, the LSCP is as follows.

$$Minimize \sum_{j \in J} f_j X_j \tag{16.1}$$

$$Subject\ to \sum_{j \in J} a_{ij} X_j \geq 1 \forall i \in I \tag{16.2}$$

$$X_j \in \{0, 1\}\ \forall j \in J \tag{16.3}$$

The decision variable x_j indicates whether base j is opened or not. The objective function (16.1) minimizes the total cost of establishing the selected facilities that has EMS vehicles stationed whilst covering all demand points. Constraints (16.2) states that each demand point must be covered by at least one opened facility. Constraints (16.3) ensure that there is no fractional part of healthcare locations. In addition to this model, another objective function which minimizes the number of located facilities is of interest. In this case, the objective function is as follows.

$$Minimize \sum_{j \in J} X_j \tag{16.4}$$

Such an objective function could be justifiable when the fixed EMS facility costs are almost equal. Daskin and Dean (2004) referred to the model minimizing the total

cost, as the set covering problem and to one minimizing the total number of located facilities, as the location set covering problem.

The above-mentioned model contains some shortcomings which are discussed in below.

- LSCM seeks the minimum amount of EMS bases required to cover all the demand points by considering unlimited amount of available stations. However, the needed resources (the cost or number of required bases) are often prohibitive. In real world problems, there is certain amount of available ambulances.
- Another disadvantage is that this model ensures that each demand point is covered just once whilst if EMS vehicles are dispatched and busy, other emergency demand points covered by the corresponding vehicles could not be covered any more.
- Furthermore, the model does not distinguish between demand nodes that relatively generate lots of demand and those that generate little demand. It is clear that the regions with more requests are of high investigation in comparison to less generated demand points.

In this regard, these concerns motivated some models to compensate for above-mentioned shortcomings. One of these models is Maximal Covering Location Problem (MCLP), developed by Church and ReVelle (1974). To formulate this model, the following notations should be denoted.

h_i	Demand at demand point i
P	Number of available EMS resources
Z_i	Binary variable, it equals to 1 if demand point i is covered, otherwise 0

Notations

Demand of node i indicates for example the number of incidents per hour or the number of residents in that area. Accordingly, the mathematical formulation is as follows:

$$Maximize \sum_{i \in I} h_i Z_i \tag{16.5}$$

$$Subject\ to\ Z_i - \sum_{j \in J} a_{ij} X_j \leq 0 \forall i \in I \tag{16.6}$$

$$\sum_{j \in J} X_j = P \tag{16.7}$$

$$X_j \in \{0, 1\} \quad \forall j \in J \tag{16.8}$$

$$Z_i \in \{0, 1\} \quad \forall i \in I \tag{16.9}$$

Objective function (16.5) maximizes the total amount of covered demands (not number of demand points). Constraints (16.6) stipulate that demand point i cannot be covered unless an EMS vehicle has been located within the relevant. Constraints (16.7) ensure that p facilities should be exactly located. Constraints (16.8) and (16.9) are standard integrality constraints. Another objective function for this model could be minimizing uncovered demand as follows.

$$Minimize \sum_{i \in I} h_i (1 - Z_i) \tag{16.10}$$

As it was mentioned before, a considerable disadvantageous is that these models do not allow multiple EMS vehicles at each base. Therefore, in real-life cases, this property results in uncovered demand points if EMS vehicles are busy for handling other calls. To deal with this shortcoming, some extended versions are developed such as Backup Coverage Problems (BACOP) (ReVelle and Hogan 1989) and Double Standard Model (DSM) (Gendreau et al. 1997).

16.2.1 Backup Coverage Problems (BACOP)

Backup coverage problems (BACOP) firstly developed by (ReVelle and Hogan 1989). BACOP models try to maximize demand points that are covered twice. The first BACOP (i.e., BACOP 1) model is similar to the MCLP but it ensures that each point is covered by at least one facility. The model's objective function maximizes the demands which are covered twice while in the second BACOP model (i.e., BACOP 2) objective function is modified rewarding demand points that are covered only once as well as those covered twice. In other words, it is combination of the demands covered once and the ones covered twice.

Backup Covering Problem 1

$$Max \sum_{i \in I} h_i \mu_i \tag{16.11}$$

$$S.t. \sum_{j \in J} X_j = P \tag{16.12}$$

$$\left(\sum_{j \in J} X_j \right) - \mu_i \geq 1 \tag{16.13}$$

$$X_j \in \{0,1\} \quad \forall j \in J \tag{16.14}$$

$$\mu_i \in \{0,1\} \quad \forall i \in I \tag{16.15}$$

Backup Covering Problem 2

$$Max\ \alpha \sum_{i \in I} h_i \mu_i + (1-\alpha) \sum_{i \in I} h_i \lambda_i \tag{16.16}$$

$$S.t. \sum_{j \in J} X_j = P \tag{16.17}$$

$$\mu_i - \lambda_i \leq 0 \tag{16.18}$$

$$\left(\sum_{j \in J} X_j \right) - \mu_i - \lambda_i \geq 1 \tag{16.19}$$

$$X_j, \mu_i, \lambda_i \in \{0,1\}\ \forall j \in J, \forall i \in I \tag{16.20}$$

In BACOP1 binary variable μ_i is defined such that it equals to 1 if and only if demand point i is covered at least twice by EMS bases. Therefore, in BACOP1, Objective function (16.11) maximizes the population covered at least twice. Constraints (16.12) states that there is only p number of EMS resources. Constraints (16.13) stipulate that each demand point is covered twice.

In BACOP2, in addition to μ_i, λ_i is defined as a binary variable for the demand points that are covered at least once. In this model, single covered demand points are taken into account by a weight factor ($\alpha \in [0,1]$). Constraints (16.18) state that demand points that are covered twice are also covered once. Constraints (16.19) imply that if demand point i is covered once then $\lambda_i = 1$; and, if the demand point is covered by more than one EMS resource then μ_i is equal to 1.

Although LSCP and MCLP are relatively simple in their formulation, they have resulted in considerable number of extensions and variants. An extension of MCLP model is the tandem equipment allocation model (TEAM) introduced by Schilling et al. (1979). In this model, two types of vehicles, primary and special vehicles, are considered. This model assumes that the coverage distance is higher for primary vehicles in comparison to secondary ones. Therefore, the maximal travel time is lower than the maximal travel time for the special ones. Daskin and Stern (1981) developed Hierarchical Objective Set Covering Model (HOSC) as a combination of LSCM and MCLP models to locate EMS vehicles in Austin (Texas). This

model simultaneously maximizes the extent of multiple coverage of demand and minimizes number of the locations. However, in real world problems, the likelihood of having two vehicles busy is low. A Double Standard Model (DSM) is a more recent model, developed by Gendreau et al. (1997). This model makes distinction between two coverage parameters (response times). It looks for solutions through which all demand points are covered at least once within response time r_1 and α percent of them must be handled within r_2. This model can be considered as an extension for BACOP1. Doerner et al. (2005) applied DSM to locate ambulances in Australia. The corresponding results demonstrate that a large proportion of population is covered in the limited amount of time and some vehicles are used more than once. Laporte et al. (2009) also applied DSM model to tackle real world problems in Belgium, Australia and Canada. Su et al. (2015) modified the DSM model by changing the objective function into minimization of operational costs and the cost of delayed services. They applied the resulted model to ambulance deployment in Shanghai, China. Backup Double Coverage model is a different backup coverage model developed by Basar et al. (2009). In this model, a demand point is covered if at least two opened bases are near to that demand zone. Church and Gerrard (2003) considered a generalization of LSCM in which multi-level coverage is needed.

16.2.2 Nonlinear Integer Programming Model for Real-Time EMS Vehicle Dispatching Model

In this subsection, the model presented by Majzoubi et al. (2012) is reviewed as a sample while rich model involving vehicle dispatching and relocating decisions. This model assigns a vehicle to a patient based on vehicle's current location and availability status of all other vehicles. Further, patients are assigned to hospitals again based on the available resources and travel times. In addition to these assignments, vehicles are relocated after completing their services to serve the next demand points. This model incorporates assumptions for some medical emergencies such as influenza outbreak. The following assumptions are made to present this model.

- The vehicle fleet is heterogeneous and can serve every patient
- All vehicles can be tracked by global positioning equipment
- Information about hospital resources are known in advance as a priori
- At each state the model is invoked for running, only the available resources are considered
- The priority of requests are known by EMS call takers

The following notations are introduced to formulate the proposed model.

Sets

I	Set of vehicles $i = 1 \ldots I$
J	Set of patients $j = 1 \ldots J$
K	Set of hospitals $k = 1 \ldots K$
S	Set of EMS stations $s = 1 \ldots S$
R	Set of census tracts $r = 1 \ldots R$
P	Set of patients' priority $p = 1, 2$
V	Set of available vehicles $V = V^P \cup V_1^H \cup V_2^H \cup V^E$
V^P	Set of vehicles en route to serve a patient
V_1^H	Set of vehicles that are able to transport high priority patients or more than one patient to the hospitals
V_2^H	Set of vehicles that are able to transport low-priority patients to the hospitals
V^E	Set of vehicles that are en route to an EMS station or they are idle
W	Set of patients or demand points $W = W^1 \cup W^2$
W^1	Set of high-priority patients
W^2	Set of low-priority patients

Parameters

T_{ij}	The estimated travel time for vehicle i to arrive at patient j's location when departing at the current time
T_{ik}	The estimated travel time for vehicle i to arrive at hospital k when departing at the current time
T_{is}	The estimated travel time for vehicle i to arrive at EMS base s when departing at the current time
T_{jk}	The estimated travel time from patient j to hospital k
μ	The average service time for serving each patient at her/his location
$PP_{pr(j)}$	The travel cost per time unit to reach patient j with priority j
IP	The travel cost per time unit of idle vehicles
A_{ij}	The travel cost for vehicle i to reach patient j; $A_{ij} = PP_{pr(j)} \times T_{ij}$
B_{ik}	The travel cost for vehicle i to arrive at hospital k; $B_{ik} = PP_{pr(j)} \times T_{ik}$
B_{ij}	The travel cost for vehicle i to reach patient j; $B_{ij} = PP_{pr(j)} \times T_{ij}$
$B_{w(i)}$	The cost of waiting for patient in a vehicle that is serving another patient; $B_{w(i)} = PP_{pr(j)} \times \mu$
B_{jk}	The cost of traveling from patient j to hospital k; $B_{jk} = PP_{pr(j)} \times T_{jk}$
C_{is}	The cost of traveling to EMS base s by vehicle i; $C_{is} = IP \times T_{is}$
D	The penalty cost incurred by serving more than one patient by one vehicle
E	The penalty cost of inefficient routing

α_p	The penalty cost of violating time window for a customer with priority p
β_r	The penalty cost of not covering census tract r by an ambulance
ρ_j	The required response time for patient j
ξ	Threshold time for reroute acceptance
ω	Threshold value for serving another patient

Variables

x_{ij}	Equals to 1 if vehicle i is dispatched to serve patient j; 0 otherwise
y_{ik}	Equals to 1 if vehicle i is dispatched to hospital k; 0 otherwise
z_{is}	Equals to 1 if vehicle i is dispatched to EMS station; 0 otherwise
π_i	Equals to 1 if ambulance i serves more than one patient; 0 otherwise
θ_i	Equals to 1 if vehicle i is not efficiently rerouted to serve more than one patient; 0 otherwise
u_j	Equals to 1 if vehicle if patient j is not served in its time window; 0 otherwise
υ_r	Equals to 1 if vehicle i is not covered by at least one ambulance; otherwise
w_i	Equals to 1 if vehicle i is not rerouted; 0 otherwise
τ_{ik}	$\tau_{ik} = T_{ik} \times y_{ik}$ if $\pi_i = 0$ and $\tau_{ik} = \sum_j T_{ik} \times y_{ik} \times x_{ij}$ if $\pi_i = 1$
ζ_{ijk}	$\zeta_{ijk} = y_{ik} \times x_{ij}$, equals to 1 if vehicle i is dispatched to serve patient j and dispatched to hospital k

It should be noted that $T_{ij,}$ T_{ik} and T_{is} are updated over the time. Using the abovementioned notations, the proposed non-linear mathematical model for the real-time vehicle dispatching (RTEMSVD) is as follows:

$$\begin{aligned} Min \sum_i \sum_j A_{ij}\, x_{ij} + \sum_{i \in V_1^H} \sum_k B_{ik} y_{ik} + \sum_{i \in V_2^H} \sum_k B_{ik} y_{ik} \left(1 - \sum_j x_{ij}\right) \\ + \sum_{i \in V_2^H} \sum_j B_{ij} x_{ij} + \sum_{i \in V_2^H} \sum_k y_{ik} \left(\sum_j x_{ij}\right) \left(\sum_j B_{jk} x_{ij} + B_{w(2)}\right) \\ + \sum_i \sum_s C_{is} z_{is} + \sum_r \beta_r \upsilon_r + \sum_p \alpha_p \sum_{j \in W^p} u_j \end{aligned} \tag{16.21}$$

$$\sum_i x_{ij} = 1 \forall j \tag{16.22}$$

$$\sum_i x_{ij} = 1 \forall i \in V^p \tag{16.23}$$

$$\sum_k y_{ik} + \sum_s z_{is} = 0 \forall i \in V^p \tag{16.24}$$

$$\sum_{i'=1} \sum_{j \in W^p} A_{ij} x^0_{i'j} - \sum_{i'=1} \sum_{j \in W^p} A_{ij} x_{ij} - \xi \geq -M w_i \forall i \in V^p \tag{16.25}$$

$$\sum_{j} \left(1 - x_{ij}\right)\ x_{ij}^{0} \leq M\left(1 - w_i\right) \forall i \in V^{p} \tag{16.26}$$

$$\sum_{k} y_{ik} = 1 \forall i \in V_1^H \tag{16.27}$$

$$\sum_{j} x_{ij} + \sum_{s} z_{is} = 0 \forall i \in V_1^H \tag{16.28}$$

$$\sum_{k} y_{ik} = 1 \forall i \in V_2^H \tag{16.29}$$

$$\sum_{j} x_{ij} \leq 1 \forall i \in V_2^H \tag{16.30}$$

$$\sum_{s} z_{is} = 0 \forall i \in V_2^H \tag{16.31}$$

$$\sum_{j} x_{ij} + \sum_{s} z_{is} = 0 \forall i \in V^{E} \tag{16.32}$$

$$\sum_{k} y_{ik} = 0 \forall i \in V^{E} \tag{16.33}$$

$$\sum_{i} x_{ij} T_{ij} - \rho_j \leq M u_j \forall j \tag{16.34}$$

$$1 - \sum_{s} \sum_{i} z_{is} \psi_{sr} \leq M \upsilon_r \forall r \tag{16.35}$$

It is clear that all the variables are binary in this model. Equation 16.21 represents the objective function that consists of eight terms including vehicles' cost to reach a patient, transporting high-priority patients, and the third through fifth terms are for low-priority patients. The sixth term is the cost of returning vehicles to EMS bases. Seventh term is a penalty for not serving patients in the defined time window and finally, the last term considers the penalty of not serving a census tract by ambulances.

Constraints (16.22) stipulate that each patient should be served by an ambulance. Constraints (16.23) through (16.26) are related to the vehicles on the way to serve patients. Constraints (16.23) state that any dispatched vehicle is assigned to a patient. Vehicles on their way can neither go to a hospital nor EMS bases; this is ensured by constraints (16.24). Constraints (16.25) and (16.26) consider route

manipulation for a vehicle if there is a reduction in travel time. Moreover, Constraint (16.27) states that vehicles, transporting high-priority patients, are assigned to a hospital and also they are not allowed to serve other patient and return to the base, presented by constraints (16.28). As Constraints (16.29) assign the low-priority patients' vehicles to hospitals and in contradiction with serving high-priority vehicles, Constraints (16.30) allow them to serve other patients. Constraints (16.31) ensure that these vehicles are not returned to EMS bases. Constraints (16.32) and (16.33) are related to idle vehicles. Constraints (16.32) allow these vehicles to remain at the same (or other) stations or serve a patient while Constraints (16.33) states that they cannot be assigned to hospitals. Response times are satisfied by Constraints (16.34) and (16.35).

16.3 EMS Operational Planning Models

As mentioned before, this chapter has focused on strategic and operational levels. In this section operational problems are investigated. Response time to reach the demand points is a critical issue for locating and relocating EMS bases and especially ambulances. Emergency medical aspects are often combined with routing and dispatching problems. Over $2.5 billion out of $5 billion of EMS expenditure is spent on transportation of patients (Sayre et al. 2001). Apart from this potential benefit, there are some other objectives that are desired to be optimized. These criteria include improving the information utilization and maximizing the productivity of services. In other words, at the operational level planning, it has to be decided which ambulance should be assigned to which request. Usually, when an emergency request occurs, the ambulance with the shortest path is dispatched (Golden et al. 2008). The contributions in ambulance operational planning are sparse and there is a little OR-based research works conducted on dispatching of vehicle fleets. Two problems, however, as ambulance dispatch and relocation are of great interest (Andersson and Värbrand 2007).

In the real world conditions, the regular transportations and emergency ones are scheduled and controlled with the same fleet of ambulances. Form another perspective, emergency transportations are required to service the emergency demand points. This kind of transportation includes dynamic elements since the parameters such as demand points, driving times and service times are uncertain and unknown at the decision making moment (Créput et al. 2011). Generally, service providers face with two different types of services based on their fleet of vehicles. First, they should cover their assigned region, as an output of location problem, in case of emergency requirements. Secondly, some regular requirements should be satisfied in terms of periodic pick-ups, delivery to patients, etc. Considering these two issues, regular transportations beside emergency ones, in one problem requires a highly robust and dynamic plan. These two problems seek different objectives. From the perspective of regular transportation, the total travel cost in terms of time or distance is minimized whilst in the case of emergency services,

providing services within a predetermined amount of time, is desired (Bruzzone and Signorile 1998). Taking these two issues into account simultaneously, the arisen problem becomes a very challenging problem. For the sake of simplicity, these two types of demands (emergency and regular transfers) can be managed independently by dividing the vehicle fleet and managing them separately. However, this attitude leads to less efficient results in comparison to the integrated approach (Kergosien et al. 2014).

Studying approaches applied to Dynamic Vehicle Routing (VRP) can provide notable insights for decision makers deciding on ambulance dispatching problems. In those VRP addressing ambulance dispatching problem, vehicles can dynamically change their routes to serve the demand points within a time window. The regular transportations are also known as "Dial a Ride Problem (DARP)" (Cordeau and Laporte 2003). This problem consists of designing vehicle routes and schedules for a set of clients to satisfy their transportation requests. In this regard, some considerations are inevitable such as vehicle capacity, maximum rid time constraint, meeting time windows, etc. Moreover, there could be different criteria like mean user ride time or total distance traveled. This problem can be employed for both static and dynamic contexts (Cordeau and Laporte 2007; Parragh et al. 2010; Attanasio et al. 2004). Beaudry et al. (2010) applied this approach into medical context. Also, Parragh (2011) studied DARP with different types of vehicles (i.e., seated, wheelchair and stretcher).

The ambulance dispatching problem is a kind of general assignment problem (Goldberg 2004). This problem is solved considering several criteria such as nearest origin, highest priority or first come first served (Haghani and Yang 2007). In other words, various dispatching rules such as Nearest Origin (NO), Highest Priority First Serve (HPFS) and First Come First Serve (FCFS) could be applied to different models. Dispatchers could have an ordered priority list for demand zones. Dispatching decisions can remarkably affect system outcome in terms of response time and survival rate. Hogan and ReVelle (1986) suggested dispatching the nearest ambulance to the demand with the highest priority. For the low-priority demands, they suggested dispatching ambulances in such a way that the demands are served in the predefined time windows. Green and Harries (1988) studied this problem and concluded that dispatching the nearest vehicle is efficient in the case of minimization of the average or total response time. Yang et al. (2004) developed a dispatching model that can be employed for EMS. Their model's objective is to minimize travel times in addition to the penalties for violating the desired time window and not covering demand points. Ibri et al. (2010) considered coverage and dispatching decisions as an integrated approach.

16.3.1 Joint Ground and Air Emergency Medical Services Coverage Model

This subsection presents the model introduced by Erdemir et al. (2010) as an operational sample model. They involved aeromedical services as well as ground ambulances in responding to trauma crashes. In this regard, they proposed a location-coverage model on the basis of both response time and total service time. This model addresses location of air and ground ambulances, and transfer points. Locating transfer points are justifiable since helicopters cannot always land at the crash sites. Therefore, it considers three options for transporting the demand requests as ground, air and joint ground-air. Two models were developed for this problem that one of them is presented here, which is a Set Covering with Backup Model (SCBM).

This model aims at covering all demand points using a combination of air and ground ambulances. As an extension to LSCP model, SCBM allows joint coverage through a combination of two facilities on a transfer point as well as exclusive coverage of two kinds of facilities. Further, other LSCP extensions consider single or backup coverage only in a direct way. To formulate the model the following notations are introduced.

Sets

M_h	Set of potential helicopter locations $h = 1 \ldots H$
M_a	Set of potential ground ambulance locations $a = 1 \ldots A$
M_r	Set of transfer point locations $r = 1 \ldots R$
N	Set of all crash paths
P	Set of all crash nodes

Parameters

C_h	Cost of locating an air ambulance
C_a	Cost of locating a ground ambulance
C_r	Cost of locating a transfer point
$A_{aj}(A_{ak})$	Equals to 1 if a ground ambulance location covers node j through path k; 0 otherwise
$A_{hj}(A_{hk})$	Equals to 1 if potential helicopter location h covers node j through path k; 0 otherwise
$A_{ahrj}(A_{ahrk})$	If potential air and ground ambulances (a and h), and transfer point location r covers node j through path k; 0 otherwise

Variables

x_a	Equals to 1 if ground ambulance is located at site a; 0 otherwise
y_h	Equals to 1 if air ambulance is located at site h; 0 otherwise
z_r	Equals to 1 if a transfer point is located at site r; 0 otherwise
u_j	Equals to 1 if node or path j is covered by at least one of the located helicopters; 0 if node or path j is covered by at least two ground ambulances and/or combinations
v_{ja}	Equals to 1 if node or path j is covered by ground ambulance a; 0 otherwise
l_{ahr}	$l_{ahr} = x_a y_h z_r$, Equals to 1 if a ground ambulance, helicopter and transfer points are located at a, h and r, respectively; 0 otherwise

Using the abovementioned notation, the proposed mathematical model is as follows:

$$Min\ Z = \sum_{a \in M_a} C_a x_a + \sum_{h \in M_h} C_h y_h + \sum_{r \in M_r} C_r z_r - \sum_{j \in N \cup P} u_j \varepsilon \tag{16.36}$$

$$s.t. \sum_{h \in M_h} A_{hj} y_h \geq u_j \forall j \in N \cup P \tag{16.37}$$

$$A_{aj} x_a + \sum_{h \in M_h} \sum_{r \in M_r} A_{ahrj} l_{ahr} \geq v_{ja} \forall j \in N \cup P, \forall a \in M_a \tag{16.38}$$

$$\sum_{a \in M_a} v_{ja} = 2\left(1 - u_j\right) \forall j \in N \cup P \tag{16.39}$$

$$x_a \geq l_{ahr} \forall a \in M_a, h \in M_h, r \in M_r \tag{16.40}$$

$$y_h \geq l_{ahr} \forall a \in M_a, h \in M_h, r \in M_r \tag{16.41}$$

$$z_r \geq l_{ahr} \forall a \in M_a, h \in M_h, r \in M_r \tag{16.42}$$

$$z_r + y_h + x_a - l_{ahr} \leq 2 \forall a \in M_a, h \in M_h, r \in M_r \tag{16.43}$$

$$z_r, y_h, x_a, l_{ahr}, u_j \& v_{ja} \in \{0, 1\} \tag{16.44}$$

Objective function (16.36) minimizes the total cost of locating ground ambulances, air ambulances and transfer points. Sum of variables u_j multiplied by a very small number ε, is subtracted from the total cost because through this subtraction, assignment of two different ground ambulances to cover node j is relaxed. Constraints (16.37–16.39) are the set covering and backup coverage constraints ensuring that all nodes are covered at least twice by ground ambulances, or once by an air ambulance and/or by a combination of air and ground ambulances. Constraints (16.40–16.43) linearized counterparts of the nonlinear terms stipulating that l_{ahr} cannot be 1 unless at least one of x_a, y_h or z is 0. In other words, this term justifies combination of air and ground ambulances used for service. So, if all x_a, y_h and z_r are 1 then all the EMS servers that form the combination are located and available.

16.4 EMS Models Under Uncertainty

The complex nature and dynamic structure of EMS systems imposes a high degree of uncertainty that could remarkably affect the overall performance of the system. Since the emergency event may happen at anytime and anyplace, uncertainty is inevitable in emergency situations. On the other hand, there are some limitations in multiple coverage models since they cannot ensure a satisfying service level. Therefore, to be of any help, we must represent the inherent problem complexities. The deterministic models do not consider the busy times of vehicle fleets. However, in real world problems, resources in terms of the total number of vehicle fleets are limited. This assumption as well as busy fractions and reliability can be covered by probabilistic EMS models. In these models, distribution functions are extracted for the arrival calls. In EMS systems, usually arrival calls are assumed to follow the Poisson processes and service is done with respect to FCFS strategy.

Daskin (1983) extended the MCLP model by considering the ambulance busy fraction. The introduced model is the Maximum Expected Covering Location Model (MEXCLP) that includes a busy fraction $\rho \in (0, 1)$ for all ambulances. In this model, three simplifying assumptions are considered as follows:

- Servers operate independently
- Servers' busy probabilities are invariant regarding the locations
- Busy probabilities are the same for all servers.

The mathematical formulation of this model is as follows.

$$Maximize \sum_j d_j \sum_{k=1}^{q} (1-\rho)\, \rho^{k-1} z_j^k \tag{16.45}$$

$$S.t. \sum_i x_i \leq p \tag{16.46}$$

$$\sum_i y_i \leq q \tag{16.47}$$

$$y_i \leq q_i x_i \forall i \tag{16.48}$$

$$\sum_i a_{ij} y_i \geq \sum_{k=1}^{q} z_j^k \forall i \tag{16.49}$$

$$x_i \,\&\, z_j^k \in \{0, 1\} \tag{16.50}$$

$$y_i \in N \tag{16.51}$$

This model overestimates coverage since independent assumptions are made. In this model, x_i is a binary variable that is 1 if a facility is set up at location i and y_i determines number of ambulances allocated to opened facility at location i. These two are limited by parameters p and q. z_j^k is a binary variable that is 1 if demand point j is covered at least by k number of ambulances. According to the base model, a_{ij} shows that if base i can cover demand point j. In this model, the objective function is to maximize the expected coverage by the term which is demand multiplied by the sum of marginal gains $(d_j(1-\rho)\rho^{k-1})$. As for the busy ambulance, ρ^k shows the probability that k ambulances are busy. Constraints (16.46) indicate that up to p emergency bases can be opened and Constraints (16.47) stipulate that total number of available ambulances is equal to q. Constraints (16.48) ensure that if a base is opened, only a limited number of ambulances can be assigned to it. Constraints (16.49) are covering limitations. Integrality constraints are shown by Constraints (16.50) and (16.51).

Simplifying assumptions of MEXCLP was later relaxed by Batta et al. (1989) through the use of Adjusted Maximum expected Covering Location Problem (AMEXCLP). As another extension, Erkut et al. (2008) incorporated busy fractions into EMS survival models entitled Maximum Expected Survival Location Problem (MEXSLP). The introduced model tends to maximize the survival probability of a patient rather than the expected coverage. In this respect, Galvao et al. (2005) developed a simulated annealing heuristic to cope with location hypercube models in order to calculate the expected coverage.

A valid strategy to incorporate the uncertainty is to determine several scenarios for uncertain parameters. For instance, uncertain demands can be estimated by a set of scenarios. In this respect, since the large number of scenarios makes the model more complex, scenario reduction methods could be useful. Nickel et al. (2016) used a sampling approach in case of large scenario numbers for ambulance location problem with stochastic demands. In this field, Bernaldi and Bruni (2009) developed

the two-stage stochastic LSCM model with uncertain demands. In this model, the first stage finds the opened bases and allocated ambulances, and the second stage assigns the demand points to the EMS bases. Van Den Berg et al. (2015) presented the linear formulation of MEXCLP with fractional coverage. The computational results show that by their formulation, run time of the linear integer programming model is considerably shorter than the corresponding non-linear one.

From another perspective, research on stochastic EMS problem is limited due to the following reasons.

- Often, quantifying the uncertainty through the probability distribution functions is too difficult because of data scarcity
- In large-scale problems such as natural disasters, the queuing models could become too complex to tackle with
- Linearization of the non-linear terms is not an easy task

To cope with these challenges, robust optimization approach can be applied in order to provide appropriate solution. This approach uses uncertainty sets rather than probability density functions or scenarios. In this method, the obtained solution is optimal against all possible realizations of the uncertain parameters within their defined sets. Robust models have been rarely applied to EMS problems. Zhang and Jiang (2014) developed a robust counterpart model for a bi-objective EMS design problem. They considered two uncertainty sources as the number of emergency calls and the maximum number of concurrent calls. Among several robust models, they applied ellipsoidal uncertainty sets for both uncertainty sources.

Another application of operations research in EMS could be through quantifying and applying resiliency concepts. Today's EMSs need to be more agile and flexible in order to deal with demand fluctuations, variations in travel times, site disruptions and so on, to quickly respond to emergency demands. In other words, disruptions in EMS could be resolved by some recovery strategies such as: back-up supplies, backup transportations modes, capacity buffers and expansions, and facility fortifications. As it was considered in this section, MEXCLP models are rather restrictive for two reasons (Ivanov et al. 2016):

- Failures or business disruption in the locations or networks are possible but not through describable probabilistic assumptions
- These models do not consider dynamics of the system that considerably impacts performance of the system.

In disruptive situations, disruption duration and recovery time considerably impact performance of the system; however it imposes dynamic dimensions to the system. In other words, the mathematical models should contain recovery functions such as differential equations to be able to describe unavailability time of the healthcare facilities and more importantly gradual recovery of capacities. Despite the rich literature on EMS models under uncertainty, most of the studies assume that the disrupted facilities are not returned to the operational mode during the planning

horizon. So, regarding the healthcare models, dynamic functions embedded in mathematical models can be helpful to develop more realistic healthcare models.

16.5 Case Study

In this section, a real facility location model case study, presented by Jia et al. (2007) is reviewed. The case study is related to large scale emergency medical services (LEMS) in case of natural disasters, terrorist attacks, and etc. local emergency response providers are designed for dealing with regular small-scale emergencies such as vehicle accidents. Location problems, however, in large scales impose special assumptions to be taken into account in order to maximize coverage and services. For instance, these kinds of services follow low rate of frequencies and overwhelmed local emergency responders that often results in national assistances. In this regard, some modifications are inevitable for definition of facility coverage. Prior to present the elaborated models, we have to make some assumptions such as facility location objectives, facility quantity and service quantity. It is noteworthy to know that redundant and dispersed requirements in facility location problem helps to enhance reliability of facilities since some emergency situations such as earthquakes affect the availability of facilities (Jia et al. 2007). Moreover, in LEMS facility locations, there are different types of coverage since occurrence of an emergency at a specific location needs more than one facility to quell the impact of the emergency. Another point is that the demand areas in LEMS should be categorized with respect to attributes such as economic importance, weather pattern, and population density, and so on.

Being mindful of modifications, Jia et al. (2007) investigates locating medical supplies for large scale services. To be more specific, protective equipment and antidotes stocks against dirty bomb attacks are to be located. In this regard, three general facility location models namely covering model, P-median model and P-center model are developed with the analyzed assumptions to formulate the location problems in large scale disasters including anthrax, dirty bomb and smallpox terrorist attacks.

According to Fig. 16.3, the considered area for locating the facilities is Lose Angeles. This area is gridded into square zones and each zone's center is considered to be aggregated demand point. Downtown, West Hollywood, LAX airport, Port of LA, Port of Long Beach, Rowland Heights and Disneyland are considered as seven demand zones.

In this figure, a number of nodes are defined as potential nodes for locating facilities. It is assumed that resources are adequate to locate just four numbers of facilities in seven potential nodes.

Developed models are employed to solve the defined problem and afterwards a comparison is made by the obtained results between the classic and developed models. The notations of these models are presented as follows.

Fig. 16.3 Los Angeles country gridded map (Adopted from Jia et al. 2007)

Sets

I	Set of possible demand points $i = 1 \ldots I$
J	Set of possible facility locations $j = 1 \ldots J$
K	Set of emergency scenarios $k = 1 \ldots K$

Variables

x_j	Equals to 1 if a facility is placed at j; 0 otherwise
z_{ij}	Equals to 1 if facility j covers demand point i; 0 otherwise
u_i	Equals to 1 demand point i is covered; 0 otherwise

Parameters

M_i	The population of demand point i
e_{ik}	The impact of demand point i under emergency scenario k
β_{ik}	The probability of emergency scenario k for demand point i
Q_i	Minimum number of facilities serving demand point i to consider demand point i as covered
p_{jk}	Service level reduction in facility j by emergency scenario k
P	The maximum available number of facilities to be located in possible locations

The mathematical formulation for dirty bomb emergency situation is as follows:

$$Maximize \sum_{i \in I} \beta_{ik} e_{ik} M_i u_i \tag{16.52}$$

$$s.t. \sum_{j \in N_i} x_j p_{jk} \geq Q_i u_i \quad \forall i \tag{16.53}$$

$$\sum_{j \in J} x_j \leq P \tag{16.54}$$

$$u_i, x_j \in \{0, 1\} \quad \forall i, j \tag{16.55}$$

This model aims at maximizing coverage under occurrence of scenario k. Constraints (16.53) ensure coverage of each demand point by required number of facilities. Constraint (16.54) enforces that the maximum number of facilities does not exceed P. Now, the P-median model in presented for Anthrax emergency problem under occurrence of scenario k. As it was mentioned earlier, in this model the objective function minimizes the total distances between demand points and their serving facilities.

$$Minimize \sum_{j \in J} \sum_{i \in I} \beta_{ik} e_{ik} M_i z_{ij} d_{ij} \tag{16.56}$$

$$s.t. \sum_{j \in J} z_{ij} p_{jk} = Q_i \quad \forall i \tag{16.57}$$

$$\sum_{j \in J} x_j \leq P \tag{16.58}$$

$$z_{ij} \leq x_j \quad \forall i, j \tag{16.59}$$

$$z_{ij}, x_j \in \{0, 1\} \quad \forall i, j \tag{16.60}$$

In this model, constraints (16.57) indicate that each demand point is covered if there are Q_i number of facilities serving it and constraints (16.59) ensures that a demand point i can be served by a facility j if the facility is already established.

In the following, the P-center model, corresponding to Small-pox emergency is presented. This model aims at minimizing the maximum distance for all demand points. As it can be seen, the only difference between the P-center and P-median model is about their objective functions, presented as objective function (16.61) and constraints (16.62).

$$Minimize \quad L \tag{16.61}$$

s.t. (16.57–16.60)

$$L \geq \frac{\sum_{j \in J} \beta_{ik} e_{ik} M_i z_{ij} d_{ij}}{Q_i} \tag{16.62}$$

Each of the proposed large-scale examples is compared with its corresponding traditional model. According to Table 16.1, in case of dirty bombs, the optimal solution of classical coverage model locates the facilities at sites 1, 4, 6 and 7. This solution ensures that 100% of population is covered by one facility while in the multiple-coverage case only 21% of the population is covered. On the other hand, the proposed model can cover 97.5% of population as single coverage and 88% as for multiple-coverage. This result implies that the developed model provides an acceptable efficiency for multiple-coverage.

For the anthrax emergency, the optimal solution, on the basis of P-center model, suggests to locate facilities at sites 1, 2, 5 and 6. As it can be seen, although for the single-coverage both models, the same distances are resulted, for multiple-coverage the objective function value of classical model is greater than the developed model. Therefore, classical model provides unbalanced solutions that could result in more economic cost value and life losses.

In the third case, the classical model suggests to locate facilities at sites 1, 2, 5 and 7 on the basis of P-median for smallpox emergency In case of single coverage models, the developed model results in larger distances (2040 versus 1740). But, if the multiple coverage is considered, the weighted total distance equals to 11,018 for traditional model while in the proposed model the weighted total distance decreases to 7,528 miles.

16.6 Future Research Directions

In this chapter, an extensive review of emergency medical services mathematical models is presented. Specifically, these models are categorized into strategic and operational level problems. Then, two rich models are studied from the recent literature and a case study is presented and investigated in order to highlight the practical value of this area. Given the current state-of-the-art literature in EMS areas, there are various avenues for further research. Among these areas we refer to the following ones:

- The proposed models can be extended to include hospital evacuation problem. When an evacuation is inevitable, emergency medical services may assist hospitals with coordinating placement and transport of patients.
- According to relevant literature, considering some real-world features such as the varying travelling speeds of vehicles, through different scenarios, could provide

Table 16.1 Comparison of the obtained results (Adopted from Jia et al. 2007)

Models	Developed models			Traditional models		
Emergency cases	Objective function		Selected sites	Objective function		
	Single coverage	Multiple coverage		Single coverage	Multiple coverage	Selected sites
Dirty bomb emergency (covering model)	Covered population: 97.5%	Covered population: 88%	1,2,3,6	Covered population: 100%	Covered population: 21%	1,4,6,7
Anthrax emergency (P-center)	Maximal weighted distance: 4*48	Maximal weighted distance: 7.5*31.4	1,2,3,7	Maximal weighted distance: 4*48	Maximal weighted distance: 7*48	1,2,5,6
Smallpox emergency (P-median)	Weighted total distance: 2040 miles	Weighted total distance: 7528 miles	1,2,3,6	Weighted total distance: 1740 miles	Weighted total distance: 11,018 miles	1,2,5,7

valuable insight for decision makers. However, EMS vehicles often do not follow all the rules and could use some exemptions.
- Following the real world problems, the current active location sites have not been included in the existed models. Therefore, it is advisable to include the current facilities in the model while making their decisions such as closing or keeping active the current facilities.
- Using bi-level programming approach including the location decisions and routing strategies, as embedded problems within another, could be another attractive subject to deal with EMS problems. In this way, global solutions rather than local ones can be obtained.
- Due to nature of inevitable uncertainties in this area, considering worst-cases could provide a remarkable vision for decision makers. Robust optimization models are helpful to reach such an understanding.

References

Andersson T, Värbrand P (2007) Decision support tools for ambulance dispatch and relocation. J Oper Res Soc 58(2):195–201

Attanasio A, Cordeau JF, Ghiani G, Laporte G (2004) Parallel tabu search heuristics for the dynamic multi-vehicle dial-a-ride problem. Parallel Comput 30(3):377–387

Başar A, Catay B, Ünlüyurt T (2009) A new model and tabu search approach for planning the emergency service stations. In: Operations Research Proceedings 2008. Berlin, Springer, pp 41–46

Batta R, Dolan JM, Krishnamurthy NN (1989) The maximal expected covering location problem: revisited. Transp Sci 23(4):277–287

Beaudry A, Laporte G, Melo T, Nickel S (2010) Dynamic transportation of patients in hospitals. OR Spectr 32(1):77–107

Beraldi P, Bruni ME (2009) A probabilistic model applied to emergency service vehicle location. Eur J Oper Res 196(1):323–331

Berman O, Drezner Z, Krass D (2011) Discrete cooperative covering problems. J Oper Res Soc 62(11):2002–2012

Brotcorne L, Laporte G, Semet F (2003) Ambulance location and relocation models. Eur J Oper Res 147(3):451–463

Bruzzone A, Signorile R (1998) Simulation and genetic algorithms for ship planning and shipyard layout. Simulation 71(2):74–83

Church RL, Gerrard RA (2003) The multi-level location set covering model. Geogr Anal 35(4):277–289

Church R, Velle CR (1974) The maximal covering location problem. Pap Reg Sci 32(1):101–118

Coppi A, Detti P, Raffaelli J (2013) A planning and routing model for patient transportation in health care. Electron Notes Discrete Math 41:125–132

Cordeau JF, Laporte G (2003) A tabu search heuristic for the static multi-vehicle dial-a-ride problem. Transp Res B Methodol 37(6):579–594

Cordeau JF, Laporte G (2007) The dial-a-ride problem: models and algorithms. Ann Oper Res 153(1):29–46

Créput JC, Koukam A, Hajjam A, Kuhn O (2011) Dynamic vehicle routing problem for medical emergency management. INTECH Open Access Publisher

Cuninghame-Green RA, Harries G (1988) Nearest-neighbour rules for emergency services. Z Oper Res 32(5):299–306

Daskin MS (1983) A maximum expected covering location model: formulation, properties and heuristic solution. Transp Sci 17(1):48–70
Daskin MS, Dean LK (2004) Location of health care facilities. In: Operations research and health care. Springer, pp 43–76
Daskin MS, Stern EH (1981) A hierarchical objective set covering model for emergency medical service vehicle deployment. Transp Sci 15(2):137–152
Doerner KF, Gutjahr WJ, Hartl RF, Karall M, Reimann M (2005) Heuristic solution of an extended double-coverage ambulance location problem for Austria. CEJOR 13(4):325–340
Erdemir ET, Batta R, Rogerson PA, Blatt A, Flanigan M (2010) Joint ground and air emergency medical services coverage models: a greedy heuristic solution approach. Eur J Oper Res 207(2):736–749
Erkut E, Ingolfsson A, Erdoğan G (2008) Ambulance location for maximum survival. Naval Res Logist (NRL) 55(1):42–58
Essen JT, Hurink JL, Nickel S, Reuter M (2013) Models for ambulance planning on the strategic and the tactical level. Beta publishing. http://doc.utwente.nl/87377
Farahani RZ, Asgari N, Heidari N, Hosseininia M, Goh M (2012) Covering problems in facility location: a review. Comput Ind Eng 62(1):368–407
Galvão RD, Chiyoshi FY, Morabito R (2005) Towards unified formulations and extensions of two classical probabilistic location models. Comput Oper Res 32(1):15–33
Gendreau M, Laporte G, Semet F (1997) Solving an ambulance location model by tabu search. Locat Sci 5(2):75–88
Goldberg JB (2004) Operations research models for the deployment of emergency services vehicles. EMS Manag J 1(1):20–39
Golden BL, Raghavan S, Wasil EA (2008) The vehicle routing problem: latest advances and new challenges: latest advances and new challenges, vol 43. Springer Science & Business Media
Gulliford M, Figueroa-Munoz J, Morgan M, Hughes D, Gibson B, Beech R, Hudson M (2002) What does' access to health care' mean? J Health Serv Res Policy 7(3):186–188
Haghani A, Yang S (2007) Real-time emergency response fleet deployment: concepts, systems, simulation & case studies. In: Dynamic fleet management. Springer, Boston, pp 133–162
Hogan K, ReVelle C (1986) Concepts and applications of backup coverage. Manag Sci 32(11):1434–1444
Institute of Medicine Committee on the Future of Emergency Care in the US Health System (I.O.M) (2006) Emergency medical services: at the crossroads. National Academy Press, Washington, DC
Ibri S, Drias H, Nourelfath M (2010) A parallel hybrid ant-tabu algorithm for integrated emergency vehicle dispatching and covering problem. Int J Innov Comput Appl 2(4):226–236
Ivanov D, Pavlov A, Dolgui A, Pavlov D, Sokolov B (2016) Disruption-driven supply chain (re)-planning and performance impact assessment with consideration of pro-active and recovery policies. Transp Res Part E: Logist Transp Rev 90:7–24
Jagtenberg CJ, Bhulai S, van der Mei RD (2015) An efficient heuristic for real-time ambulance redeployment. Oper Res Health Care 4:27–35
Jia H, Ordóñez F, Dessouky M (2007) A modeling framework for facility location of medical services for large-scale emergencies. IIE Trans 39(1):41–55
Karasakal O, Karasakal EK (2004) A maximal covering location model in the presence of partial coverage. Comput Oper Res 31(9):1515–1526
Kergosien Y, Gendreau M, Ruiz A, Soriano P (2014) Managing a fleet of ambulances to respond to emergency and transfer patient transportation demands. In: Proceedings of the International Conference on Health Care Systems Engineering. Springer International Publishing, pp 303–315
Kerkkamp RBO (2014) Facility location models in emergency medical service: robustness and approximations (Doctoral dissertation, TU Delft, Delft University of Technology)
Knight VA, Harper PR, Smith L (2012) Ambulance allocation for maximal survival with heterogeneous outcome measures. Omega 40(6):918–926

Lam SSW, Nguyen FNHL, Ng YY, Lee VPX, Wong TH, Fook-Chong SMC, Ong MEH (2015) Factors affecting the ambulance response times of trauma incidents in Singapore. Accid Anal Prev 82:27–35

Laporte G, Louveaux FV, Semet F, Thirion A (2009) Application of the double standard model for ambulance location. In: Innovations in distribution logistics. Springer, Berlin/Heidelberg, pp 235–249

Li X, Zhao Z, Zhu X, Wyatt T (2011) Covering models and optimization techniques for emergency response facility location and planning: a review. Math Meth Oper Res 74(3):281–310

Looije D (2013) Optimal ambulance locations in the Netherlands (Doctoral dissertation, Delft University of Technology)

Majzoubi F, Bai L, Heragu SS (2012) An optimization approach for dispatching and relocating EMS vehicles. IIE Trans Healthc Syst Eng 2(3):211–223

Nickel S, Reuter-Oppermann M, Saldanha-da-Gama F (2016) Ambulance location under stochastic demand: a sampling approach. Oper Res Health Care 8:24–32

Parragh SN (2011) Introducing heterogeneous users and vehicles into models and algorithms for the dial-a-ride problem. Trans Res Part C: Emerg Technol 19(5):912–930

Parragh SN, Doerner KF, Hartl RF (2010) Variable neighborhood search for the dial-a-ride problem. Comput Oper Res 37(6):1129–1138

ReVelle C, Hogan K (1989) The maximum availability location problem. Transp Sci 23(3):192–200

Sayre MR, White LJ, Brown LH, McHenry SD (2001) National EMS research agenda. Prehospital Emerg Care: Off J Natl Assoc EMS Phys Natl Assoc State EMS Dir 6(3 Suppl):S1–43

Schilling D, Elzinga DJ, Cohon J, Church R, ReVelle C (1979) The TEAM/FLEET models for simultaneous facility and equipment siting. Transp Sci 13(2):163–175

Schmid V (2012) Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. Eur J Oper Res 219(3):611–621

Schmid V, Doerner KF (2010) Ambulance location and relocation problems with time-dependent travel times. Eur J Oper Res 207(3):1293–1303

Soriya G, Colwell CB (2012) Emergency medical services. In: Encyclopedia of intensive care medicine. Springer, Berlin Heidelberg, pp 839–840

Su Q, Luo Q, Huang SH (2015) Cost-effective analyses for emergency medical services deployment: a case study in Shanghai. Int J Prod Econ 163:112–123

Toregas C, Swain R, ReVelle C, Bergman L (1971) The location of emergency service facilities. Oper Res 19(6):1363–1373

Van Buuren M (2011) TIFAR modeling package for the evaluation of emergency medical services: with EMS modeling results for the Amsterdam area (Doctoral dissertation, TU Delft, Delft University of Technology)

Van den Berg PL, Kommer GJ, Zuzáková B (2015) Linear formulation for the maximum expected coverage location model with fractional coverage. Oper Res Health Care 8:33–41

Yang S, Hamedi M, Haghani A (2004) Integrated approach for emergency medical service location and assignment problem. Transp Res Rec: J Transp Res Board 1882:184–192

Zhang ZH, Jiang H (2014) A robust counterpart approach to the bi-objective emergency medical service design problem. Appl Math Model 38(3):1033–1040

Chapter 17
Health Informatics

I. Burak Parlak and A. Cağrı Tolga

17.1 Introduction

Over last decades, hardware and software implementations stimulated new research areas in medicine by inspiring new algorithms through the applications. The multidirectional cooperation between engineering and medicine generated the use of bio-inspired models in engineering and the mathematical routines in medicine, respectively. Then, the integration of computational tools into the medicine builds up new foundations and algorithmic designs for bioengineering and medical informatics. Those applied disciplines involved in miniaturization and nanotechnology are inherently still in progress due to the restrictions of undiscovered effects of physiological systems, genetics, drugs or pharmaceutics. The cutting edge in all these applications is to adapt current algorithms to handle big data within optimal computation time.

In this chapter, crucial optimization and medical decision aspects are considered within different levels of medical applications. Figure 17.1 summarizes a general point of view for medical informatics. Different applications in medicine could be linked onto computational tools to resolve unconsidered effects of diseases. A drug delivery system based on nano particles, augmented reality applications through medical robotics, cellular imaging and visualization tools in neuroscience are some of the illustrative examples within the use of medical informatics. Historically, the use of soft-computing techniques has started within the instrumentation tools and

I.B. Parlak
Department of Computer Engineering, Galatasaray University, Istanbul, Turkey
e-mail: bparlak@gsu.edu.tr

A.C. Tolga (✉)
Department of Industrial Engineering, Galatasaray University, Istanbul, Turkey
e-mail: ctolga@gsu.edu.tr

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_17

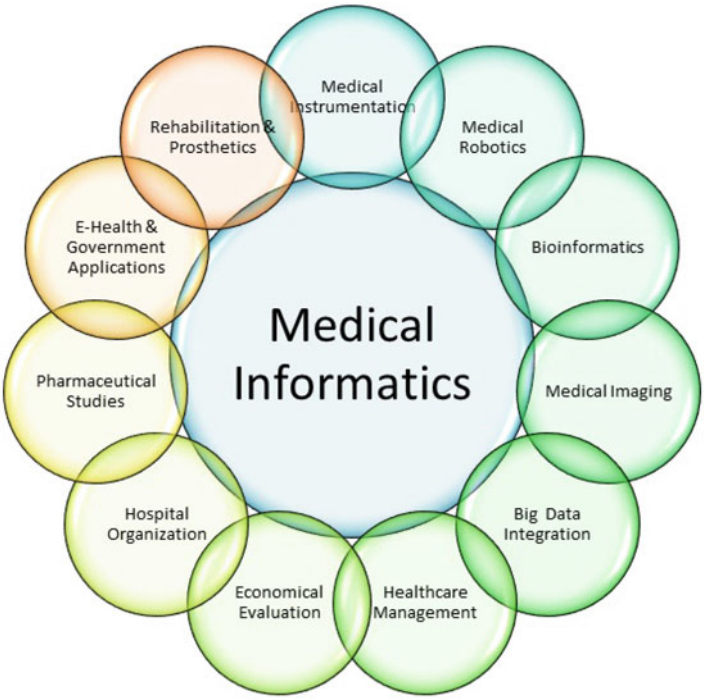

Fig. 17.1 A general representation of medical informatics and related disciplines

the medical imaging. The micro controllers and low level programming languages have played a major role in the development of medical informatics to guide the medical diagnosis and the patient survey. Nowadays, the use of intelligent markers in cellular imaging or the nanofibers in tissue engineering allows us to use computing techniques in micro/nano medicine. Therefore, the limitation of the chapter leads us to make a choice of optimization techniques and a selection among medical applications in order to cover medical computing challenges from micro to macro levels.

The cutting-edge topics in bioinformatics was triggered regarding Human Genome Project over the last decade. The multiobjective optimization in computational biology has several applications in genomics. Diagnosing and curing several diseases were transferred into genomic levels where the engineering processes and algorithmic approaches were reinterpreted. Then, genomic optimization problem was integrated into this field. Nowadays, genome databases allow us to use data mining, classification, and especially machine learning tools in biological sequences. Phylogenetic inference to assess evolutionary trees between species, gene regulatory networks to understand the mechanism of nucleic acids and proteins, sequence alignment to highlight the similarities of nucleic acids and protein analysis to predict the functionality of macromolecules are the major research fields of bioinformatics (Handl et al. 2007). Sequence analysis of virology

is an important problem to predict pandemic and epidemic diseases and also to develop prospective drugs. The problem could be analyzed with two major approaches; global and local alignment techniques.

Medical image segmentation is a fundamental problem of medical image processing. Initially, medical image analysis was performed with manual or semi-automatic techniques in radiology. Image contours, regions and contrasts are crucial to determine potential diseases. In modern digital radiology, this step could be performed with automatic techniques where several algorithms are in use. Here, the energy minimization problem is presented within the medical image segmentation. Variational methods are fundamental tools in mathematics. Its application onto radiology speeds up the image analysis in terms of preprocessing. Briefly, the similarity measure in 2D/3D images is calculated with the displacement vectors based on a cost function. Even though there are several similarity metrics, Mumford-Shah formulation (Tsai et al. 2001), their approximations, region based and edge based active contours are characterized as robust approaches. The edge set parametrized as a geometrical object is used in the segmentation problem where the optimization procedure minimizes the quantity of shape displacement fields in order to reach the target area.

Another optimization problem in modern radiology is the image registration. Similarly, the problem is considered as an essential step in image analysis where the patient or internal organs would cause deformations and the physician may not properly visualize the image and take a decision. The aim of the problem is briefly to map corresponding features between source and target images. Therefore, the alignment process may eliminate or reduce the misalignments that occurred during 2D/3D image acquisition. This mapping function may depend on the scanner configuration, the patient characteristic or the application area (e.g cardiac or cranial imaging). The quality of superimposed images is assessed by distance minimization based on linear or non-linear optimization procedures. Target images are obtained using a geometrical transformation where the intensity, shape, contour, statistical features of images are used. A robust registration algorithm Iterative Closest Point (ICP) based on Levenberg-Marquart optimization was introduced for intermodal applications (Castellani and Bartoli 2012; Fitzgibbon 2003). This hybrid application allows us to satisfy the preprocessing steps in automatic medical assessment and to guide the temporal medical survey of patients.

Healthcare information systems are commonly in use to increase the efficiency of hospital organization, to share the knowledge network between medical departments, and to set soft computing platform in order to analyze all information about the hospital. In this chapter, we studied this problem within the scope of strategic reorganization to set up a private hospital. The key concepts of hospital engineering are revised using Multi Criteria Decision Making (MCDM) approach to identify the needs of a complete hospital and to clarify the relationships among them.

Finally, a challenging medical decision problem in radiology; the selection of medical imaging devices is studied. The problem is limited with conventional radiology because general hospital investments are mostly related with the imaging equipments and interventional techniques are not commonly in use for all hospitals.

The new branches, remote soft-computing architectures, telemedicine, distant medical diagnosis, patient survey are also combined with the technological properties of these devices. The increase of imaging quality imposes new costs for the investment and corresponding staff. Therefore, a detailed study of medical device selection was performed using VIKOR (VIseKriterijumska Optimizacija i kompromisno Resenje) approach.

In a nutshell, the second section will revise sequence analysis problem in genomics and virology. The third part will detail the segmentation problem using active contours. The fourth section will introduce the image registration problem in multimodal image analysis. The fifth part will detail a strategic reorganization problem as a hospital engineering problem. Moreover, the next section will clarify how MCDM tools could be used in a medical problem where imaging devices are observed. We conclude this chapter with an overview of current and future trends in health informatics.

17.2 Optimizational Computing in Bioinformatics

The Human Genome Project created a new bridge to fill the gaps between life sciences and engineering at the level of chromosomes and nucleic acids. Over last decades several bioinformatic tools were interpreted to design soft-computing techniques. Biological data mining and bioinformatic databases rendered meaningful information about the use of genomics and proteomics. The extraction and the analysis of sequences were performed within different platforms. A brief introduction of related terminology would highlight the basics of a biological database (Fulekar 2009);

- Proteomics: The functionality of a protein sequence can be analyzed within the protein databases. The meaningful patterns are highlighted regarding the user query.
- Similarity analysis: The similarity measure between multiple sequences can be analyzed within a pattern quest in a database. The similarity identification and the distance between those structures are characterized with different mathematical metrics.
- Structural analysis: The morphological properties of a protein's shape is mandatory to perform a comparison or to visualize the searched effect. Therefore, the morphology and the structure are observed in details regarding the requested pattern.
- Sequence analysis: A detailed analysis of user query is performed using evolutionary biology where genetic operators like crossover and mutation are taken into account.

The section will start with the review of pairwise sequence alignment. We will apply an adapted form of dynamic programming on H1N1 influenza sequences of 2009 pandemic retrieved from GenBank database (http://www.ncbi.nlm.nih.gov/genbank/). Genomic information based on nucleic acids (DNA, RNA) or protein

sequences could be studied by the maximization of similarity score between them. Then, a quantification score between genomic quests would be performed and the distance between different genomic types would be inferred.

Even if this search process could be studied with several approaches, we will take into account two main techniques; global alignment based on Needleman-Wunch (NM) algorithm and local alignment using Smith-Waterman (SW) algorithm. Smith et al. (2009) studied the origins and evolutionary genomics of H1N1 influenza based on phylogenetic analyses. Moreover, Gallaher (2009) discussed the changes of Hemagglutinin for 2009 H1N1 influenza and vaccine strategies. Koppstein et al. (2015) performed sequence analysis to reveal the underlying mechanism of viral transcription for H1N1 influenza.

In order to compare sequences, the initial operators should be revised. The substitution, the insertion and the deletion are considered as the initial operators to compute a total score of the alignment problem. We may use the following function to compute a score value;

$$sim(s_1, s_2) = \begin{cases} +1 & \mathrm{s}_1 = \mathrm{s}_2 \\ -1 & \mathrm{s}_1 \neq \mathrm{s}_2 \\ 0 & \text{otherwise} \end{cases} \tag{17.1}$$

In order to find the total alignment θ, we perform this procedure along the sequence. Then, $\theta(s) = \sum_i sim(s_{1_i}, s_{2_i})$.

If s_1 and s_2 are two sequences, where $s_1 = \texttt{CAGT}$ and $s_2 = \texttt{TGA}$, we may calculate the alignment value θ as follows;

$$\theta = \begin{matrix} \texttt{C A G T} \\ \texttt{T G} - \texttt{A} \end{matrix}$$

Then, the score value is

$$\theta(s) = sim(\texttt{C}, \texttt{T}) + sim(\texttt{A}, \texttt{G}) + sim(\texttt{G}, -) + sim(\texttt{T}, \texttt{A}) = -1 - 1 + 0 - 1 = -3$$

This alignment procedure might be generalized into a matrix form. Here, we will integrate a global alignment problem for two genomic sequences using NM algorithm which is based on a dynamic programming. The basic idea is to create optimal alignment between two sequences s_1 and s_2 using the subsequences.

Let s_1 and s_2 be two sequences, if we want to perform an optimal alignment of a length i between them, we will start with the initial step where we may have three different cases;

1. $\begin{matrix} s_{1_i} \\ | \\ s_{2_i} \end{matrix}$; when there are elements to compare on both sequences

2. $\begin{matrix} s_{1_i} \\ | \\ - \end{matrix}$; when the second sequence has a gap

3. $\begin{matrix} - \\ | \\ s_{2_i} \end{matrix}$; when the first sequence has a gap

Then, we may build up a linear representation within a matrix form M to resolve these cases. A score $M_{i,j}$ sets the score of an optimal alignment between $s_{1_{(1\ldots i)}}$ and $s_{2_{(1\ldots j)}}$.

$$
\mathrm{M} = \begin{array}{c|c}
 & 0\ 1\ \ldots \quad i-1 \qquad i \quad \ldots \quad n \\
\hline
0 & \\
1 & \\
\vdots & \\
j-1 & \quad M_{i-1,j-1} \quad M_{i,j-1} \\
 & \qquad\qquad \searrow \ \downarrow \\
j & \quad M_{i-1,j} \longrightarrow \ M_{i,j} \\
\vdots & \\
m & \qquad\qquad\qquad\qquad M_{m,n}
\end{array}
$$

$M_{m,n}$; bottom right corner of the matrix is denoted as θ; the score of optimal alignment. Then, for all cell values, we will use its three neighbourhoods to update them and to reach the right corner.

$$
\begin{array}{cc}
M_{i-1,j-1} & M_{i,j-1} \\
 & \searrow \ \downarrow \\
M_{i-1,j} & \longrightarrow \ M_{i,j}
\end{array}
$$

$$
M_{i,j} = \max \begin{cases} M_{i-1,j-1} + sim(s_{1_i}, s_{2_j}) \\ M_{i-1,j} + sim(s_{1_i}, -) \\ M_{i,j-1} + sim(-, s_{2_j}) \end{cases} \tag{17.2}
$$

In order to cope with the boundary conditions; For $j = 0$, the value of $M_{i,0}$ is not initialized, then it is set as $M_{i,0} = \sum_i sim(s_{1_i}, -)$. Furthermore, for $i = 0$; the left column cells $M_{0,j}$ are set as $M_{0,j} = \sum_j sim(-, s_{2_j})$.

This algorithm is called a global alignment because we are aligning one sequence against another sequence start to end. There are other similarity techniques like local alignment (such as Smith Waterman), multiple alignment (such as CLUSTAL and CLUSTALW2) and fast alignment (BLAST-Basic Local Alignment Search Tool).

Global alignment might be considered as a robust technique in sequence alignment process related to annual pandemic where the percentage of similar patterns would be higher. However, it is remarkable that a virus like H1N1 might evolve quickly and the similarity in a pandemic might differ easily.

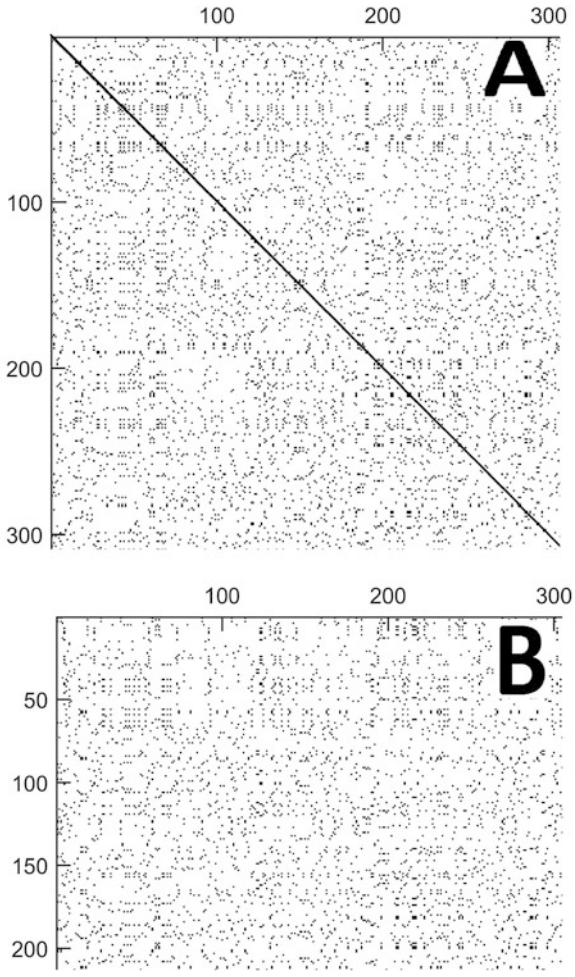

Fig. 17.2 Sequence plot between two H1N1 Influenza sequences (**a**) High correlation: the diagonal of the matrix indicates the edit distance. (**b**) Low correlation

The results denoted on Figs. 17.2 and 17.3 represent the high and low similar cases for the influenza pandemics of 2009. It is crucial to note that if there exists a linear line on the diagonal of similarity matrix we may infer that those sequences are quite similar and highly correlated. Therefore a global alignment technique would be satisfactory. On the other hand, if there is not a linear correspondence between them it means that the correlation is low and a local alignment would be performed.

The high and low correlated sequences illustrated in Fig. 17.2 characterize the initialization of alignment process. The diagonal line on the alignment matrix is useful to see if the process is adapted to perform a global or a local alignment. Therefore, the analysis in Fig. 17.3 depicts the corresponding results for the global alignment. It is remarkable that H1N1 may differ quickly even the sequences belong to the same location and/or the same year.

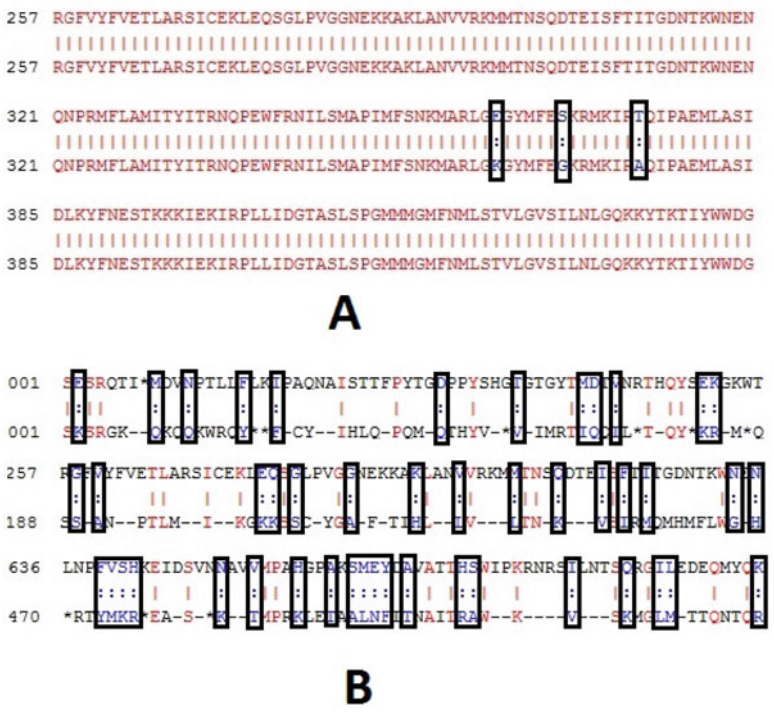

Fig. 17.3 Sequence transition between two H1N1 Influenza sequences (**a**) High correlation: small number of *black boxes*. (**b**) Low correlation: large number of *black boxes*

The comparison of nucleic acids and protein sequences requires an access to biological databases to perform a search query. The result of the comparison is to determine its relationship to other nucleic acids or proteins. Therefore, proteomics and sequence analysis in virology could be transformed on the representation of strings of characters where the corresponding query is characterized and mapped on other patterns.

Secondly, we will introduce a local alignment using SM algorithm. In sequence analysis of virology SM, FASTA, and BLAST are popular techniques to see the similarity between viruses. BLAST and FASTA may not give not accurate results but they are faster than SM algorithm Therefore, we preferred SM to see reliable local alignment results for 2009 pandemics.

This time we will consider finding the best alignment between the subsequences s_1 and s_2. There are few modifications to the global alignment algorithm so it will fit to the problem of finding the best local alignment.

In order to apply local technique, start and end states within sequence strings should be noted. In this case, starting and closing gaps might be ignored. Briefly, SM technique is as follows;

Fig. 17.4 Local alignment using Smith-Waterman algorithm

- Instead of the maximal score obtained at the bottom right corner of the global matrix M, decompose the alignment into local matrices.
- The alignment could be performed within local matrices.
- The best alignment score might be calculated in local matrices.
- The summary of local alignment procedure is based on the following operator where the negative values are not taken into account;

$$M_{i,j} = \max \begin{cases} M_{i-1,j-1} + sim(s_{1_i}, s_{2_j}) \\ M_{i-1,j} + sim(s_{1_i}, -) \\ M_{i,j-1} + sim(-, s_{2_j}) \\ 0 \end{cases} \tag{17.3}$$

- A new value 0 represents the start for a new alignment in a local matrix any starting point.
- If the alignment score to the end of local matrix has a negative value, we might search a new alignment rather than extending this local alignment.

In order to restore the best local alignment one can start at the highest matrix score cell and start backtracking using the father pointers back to a 0 value cell. This path corresponds to the best local alignment (Fig. 17.4).

17.3 Variational Methods in Image Segmentation

Image segmentation is one of the fundamental problems in medical image analysis. The aim is to divide an acquired image or volume into separate partitions within the medical interest. These partitions are characterized with their region boundaries, their contours, their centroids and other morphological properties. Recently, methods based on Partial Differential Equations (PDE) have been found effective in order to perform this extraction process. Even though, the extracted shape depends on the mathematical background, current tools allow us to optimize this procedure.

Variational methods are briefly categorized within the shape optimization where an energy function depending on the pixels/voxels deforms the curve with respect to the energy minimization. Therefore, the final curve gives a boundary where the extracted area is located. Gooya et al. (2008) used variational methods to perform the segmentation in angiographic vascular images. Li et al. (2011) developed a level set approach in magnetic resonance segmentation and bias correction. Active contours or snakes are considered as a subclass of variational methods. Ho et al.

(2002) presented a level-set snakes in the identification of brain tumors based on 3D images. Even though, their initialization and convergence in medical contours could reduce the sensitivity, the major advantage of them is their robustness for soft-tissue extraction such as tumor area analysis in prostate, liver or brain.

In this section, Kass (1998)'s model; a common active contour model (Kass et al. 1988) will be briefly presented. A traditional snake or active contour, C is formulated with $C(s) = [x(s), y(s)], s \in [0, 1]$.

$$\begin{aligned} C : [0, 1] &\rightarrow \mathbb{R}^2 \\ s \quad &\rightarrow C(s) = (x(s), y(s)) \end{aligned} \tag{17.4}$$

It moves spatially on a grayscale image $A(x, y)$ to minimize the energy functional E. In order to express this function;

$$\begin{aligned} x' \equiv \frac{dx(s)}{ds}, x'' \equiv \frac{d^2x(s)}{ds^2} \\ y' \equiv \frac{dy(s)}{ds}, y'' \equiv \frac{d^2y(s)}{ds^2} \end{aligned} \tag{17.5}$$

should be defined.Then, the initial conditions are expressed $x(0) = x(1)$, $x'(0) = x'(1)$, $x''(0) = x''(1)$, $y(0) = y(1)$, $y'(0) = y'(1)$, $y''(0) = y''(1)$ to obtain the following general formulation.

$$E_{total}(v) = E_{int}(v) + E_{ext}(v) \tag{17.6}$$

Then, we may write

$$E = \int_0^1 \frac{1}{2}(\alpha * |C'(s)|^2 + \beta * |C''(s)|^2) + E_{ext}(C(s))ds \tag{17.7}$$

where α and β denote the control parameters of the *elasticity* and the *stiffness*, respectively. The derivatives; $C'(s)$ and $C''(s)$ are calculated with respect to s. Furthermore, the external energy function E_{ext} is expressed regarding the image features on the region as follows;

$$E_{ext}(x, y) = -|\nabla(\mathcal{N}_{\mu,\sigma}(x, y) * A(x, y))| \tag{17.8}$$

Here, $\mathcal{N}_{\mu,\sigma}$ denotes a 2D Gaussian function with the mean μ and the standard deviation σ. ∇ is a gradient operator to complete the computation. The Gaussian function will smooth the boundaries and larger σs will increase the noise on the contours.

Then, the snake should satisfy the following Euler-Lagrange equation;

$$\alpha * C''(s) - \beta * C''''(s) - \nabla E_{ext} = 0 \tag{17.9}$$

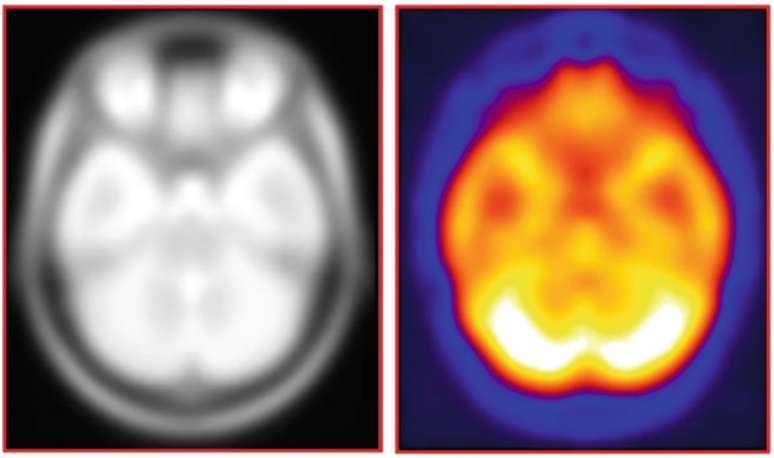

Fig. 17.5 Initial slices. (**a**) PD-MRI, (**b**) PET

which is written equivalently as;

$$F_{int} + F_{ext}^{(p)} = 0 \tag{17.10}$$

where

$$F_{int} = \alpha * C''(s) - \beta * C''''(s) \tag{17.11}$$

$$F_{ext}^{(p)} = -\nabla E_{ext} \tag{17.12}$$

In order to implement the snake algorithm, we will use two different 3D image modalities; proton density (PD) Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) which will be used also in the next section for image registration. The initial slices are represented in Fig. 17.5. We should apply a discrete representation of snakes onto the slices. Then;

The initial curve which is sketched as a red frame in Fig. 17.5 on both modalities is expressed by K_n control points:

$$(x_i, y_i) = (x(c_i), y(c_i)) \quad \forall i \in \{0, K_n - 1\} \tag{17.13}$$

The curve C which is parametrized by s is discretized where $s_i = c_i$. Therefore, we may evaluate the discrete differences to obtain the energy function and to find the optimal solution by the gradient descent algorithm;

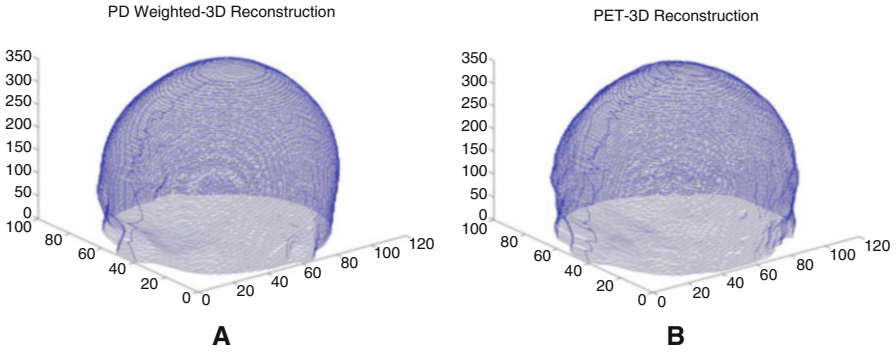

Fig. 17.6 Scalp segmentation results. (**a**) 3D PD-MRI, (**b**) PET images

$$E_{int}(c) = h\sum_{i=0}^{K_n}\left[\alpha\frac{1}{2h^2}((x_{i+1}-x_i)^2+(y_{i+1}-y_i)^2)\right.$$
$$\left.+\,\beta\frac{1}{2h^4}((x_{i-1}-2x_i+x_{i+1})^2+(y_{i-1}-2y_i+y_{i+1})^2)\right] \tag{17.14}$$

Consequently, the initial red frame reaches the skull for all slices. These slices are reconstructed to create the 3D volumes. The final segmentation results are represented in Fig. 17.6. Therefore, we may remark that variational methods could achieve an essential preprocessing step on cranial segmentation.

17.4 Image Registration

In medical imaging, patient and disease survey requires periodical acquisitions. These images should be preprocessed in order to bring the same coordinates onto the same pixels/voxels. For this purpose, source and target images should be registered. This geometrical problem has many constraints and drawbacks. The anatomical landmarks, medical acquisition problems, artefacts and image quality may reduce the sensitivity of registration. Furthermore, no registration is needed between two images if the same objects do not change shape spatially and/or their relationship to the image coordinate system does not alter.

In this section, Iterative Closest Point Algorithm (ICP); a registration heuristics from both source and target images to bring them in a common coordinate system (Fitzgibbon 2003; Besl 1992) will be presented. ICP has been noted as a robust tool in the preprocessing of medical acquisitions for skull, abdominal and chest imaging. Its extended forms include different transformation phases, the integration of intensity or curvature information within the minimization function. Maurer

et al. (1996) applied weighted geometrical features to compute the registration map in X-ray computed tomography (CT) and T2-weighted magnetic resonance volume in cranial imaging. Su et al. (2009) applied 3D ICP registration process for stereoscopic videos in kidney imaging during robot-assisted nephrectomy. Duan et al. (2104) developed a skull identification system to measure the correlation between skull and face using ICP in CT images. During this section, we will start with the initial stage of ICP in order to extend it with Levenberg-Marquart optimization tool.

The discrete formulation of medical image registration; N_s and N_t represent the number of points for the source $S = s_i$, $i = 1, \cdots, N_s$ and the target $T = t_j$, $j = 1, \cdots, N_t$ respectively. Here, s_i and t_j are 3-D vectors: voxel coordinates.

The correspondence between source and target points will be evaluated and the optimal registration vector will be calculated using these coordinates. The following step is to set the translation and rotation parameters. This definition will lead to the optimum registration between two sets. The application of this transformation will enable to obtain the mapping. This process is applied iteratively until a certain dissimilarity measure becomes smaller than a certain value.

ICP algorithm registers sets S, T by evaluating for each point s_i in S, its corresponding point in T, i.e., the point in T with the minimum distance. The evaluation of the distance between a point and a data set is denoted by $d(s_i, T)$:

$$d(s_i, T) = \min_{j \in \{1, \cdots, N_t\}} d(s_i, t_j) \tag{17.15}$$

where $d(s_i, t_j)$ is the Euclidean distance of the two points. It is repeated for all points in S so that, for each point s_i, the closest point t_j in T is found. This results in the construction of the set Y of closest points, a process that is denoted by the operator C:

$$Y = C(S, T) \qquad where \qquad N_y = N_s \tag{17.16}$$

After the evaluation of the correspondence between two point sets, ICP proceeds to the determination of the translation and rotation parameters. This is performed using *quaternions*. A unit quaternion is a linear tool to represent 3D rotations It is a four-dimensional vector $q_R = \begin{bmatrix} q_0 & q_1 & q_2 & q_3 \end{bmatrix}^T$ such that $q_0 \geq 0$ and $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. Rotation by an angle θ around a unit vector $[x_1 \; x_2 \; x_3]^T$ can be formulated by the following quaternion:

$$q_R = \begin{bmatrix} \cos\frac{\theta}{2} & x_1 \sin\frac{\theta}{2} & x_2 \sin\frac{\theta}{2} & x_3 \sin\frac{\theta}{2} \end{bmatrix}^T \tag{17.17}$$

The rotation matrix $\mathbf{R}$ that corresponds to a certain quaternion is the following:

$$R(q_R) = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 + q_2^2 - q_1^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 + q_3^2 - q_1^2 - q_2^2 \end{bmatrix} \tag{17.18}$$

Furthermore, a 3D translation is expressed by the vector; $q_T = \begin{bmatrix} q_4 \; q_5 \; q_6 \end{bmatrix}^T$. The two vectors are merged to obtain the registration vector.

$$q = \begin{bmatrix} q_R \\ q_T \end{bmatrix} = \begin{bmatrix} q_0 \; q_1 \; q_2 \; q_3 \; q_4 \; q_5 \; q_6 \end{bmatrix}^T \tag{17.19}$$

Then, the optimal registration vector is found using the minimization of the mean square error function:

$$f(q) = \frac{1}{N_s} \sum_{i=1}^{N_s} \| y_i - R(q_R) s_i - q_T \|^2 \tag{17.20}$$

In order to calculate the optimal q_R, q_T, the centers of the mass μ_s, μ_y of point sets S, Y, are evaluated, respectively:

$$\begin{aligned} \mu_S &= \tfrac{1}{N_s} \textstyle\sum_{i=1}^{N_s} s_i \\ \mu_Y &= \tfrac{1}{N_y} \textstyle\sum_{i=1}^{N_y} y_i \end{aligned} \tag{17.21}$$

The cross-covariance matrix of S, Y is calculated as follows:

$$\sum_{s,y} = \frac{1}{N_s} \sum_{i=1}^{N_s} \left[(s_i - \mu_s)(y_i - \mu_y)^T \right] \tag{17.22}$$

The column vector Δ is written;

$$\Delta = \begin{bmatrix} A_{23} \; A_{31} \; A_{12} \end{bmatrix}^T \tag{17.23}$$

where A_{ij} are the i, j elements of the matrix $(\sum_{sy} - \sum_{sy}^T)$. Finally, a 4×4 matrix of the following form is generated:

$$Q(\sum_{sy}) = \begin{bmatrix} tr(\sum_{sy}) & \Delta^T \\ \Delta & \sum_{sy} + \sum_{sy}^T - tr(\sum_{sy}) I_3 \end{bmatrix} \tag{17.24}$$

where $tr()$ is the trace of a matrix, which is the sum of its diagonal elements, and I_3 is the 3×3 identity matrix. The unit eigenvector that corresponds to the maximum eigenvalue of $Q(\sum_{sy})$ is the optimal rotation quaternion q_R. The optimal translation vector is chosen to be

$$q_T = \mu_y - R(q_R) \mu_s \tag{17.25}$$

The parameters evaluated in the previous step are then used to translate and rotate the data set S;

$$S = q(S) \tag{17.26}$$

where $q(S)$ denotes transformation of set S using the registration vector q. For the first iteration of the algorithm, the data set is initialized using the initial point set $S_0 = S$, whereas the registration vector q is initialized as follows:

$$q_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{17.27}$$

Then, ICP proceeds as follows :

1. The closest points are computed using;

$$Y_k = C(S_k, T) \tag{17.28}$$

2. The optimal registration vector q_k is evaluated.
3. The data set S is translated and rotated using the registration vector evaluated in the previous step:

$$S_{k+1} = q_k(S_k) \tag{17.29}$$

4. Use the Levenberg-Marquart algorithm to calculate the mean squared error. If it is below a certain threshold then the algorithm stops. Otherwise, steps 1 through 4 are repeated using S_{k+1}.

Levenberg-Marquart (LM) optimization is a tool to minimize the error between measured and expected data. Then, it may be applied into image registration problem where source and target points differ. This procedure takes into account the sum of squared residuals (the error function); E and their derivatives are computed. It combines the gradient descent and Gauss-Newton approaches to function minimization (Fitzgibbon 2003).

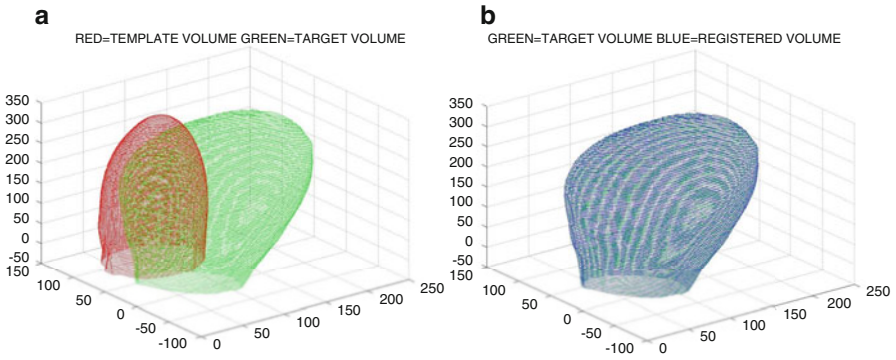

Fig. 17.7 Intermodal scalp registration using ICP Levenberg-Marquart optimization: (**a**) source PD-MRI (*red volume at left*) registered to target PET (*green volume at right*). (**b**) Registered image and target PET image

LM-ICP procedure was applied on two different image modalities; PD-MRI and PET which were already constructed in the previous section. As an intermodal registration PD-MRI was warped on PET volume and final mapping was found through the minimization of LM. Figure 17.7 represents the initial and the final case where the registered volume coincides on PET image.

17.5 Strategic Reorganization Planning Among Healthcare System Units

A reorganization implies restructuring of any foundation's legal, ownership, operational, or other structures to make the organization more profitable or renew it for present needs. In addition financial reorganization comprehends the restatement of assets and liabilities. Reorganization is an interference to prolong the life of a company facing troubles through special arrangements and restructuring for the purpose of minimizing the possibility of past situations reoccurred. The nature of word "strategy" looks to the future. Therefore, strategic reorganization process needs to be capable of farsightedness without an occurrence of trouble or profit reduction. Organizations sometimes might have abundance of cash, however in these cases they should think about the future threats and take precautions against them. At these times reorganizing some departments may make snowball effect to allow too much cash.

Healthcare systems engineering which integrates the healthcare and engineering disciplines provides solution processes to effectiveness, safety, and efficiency in healthcare. Growing complexities in health care drove the researchers to look from the system though engineering point of view. However this perspective needs to be patient-centered focus while it should bear in mind the organization, structure, function, productivity and delivery of services in health care. Many criteria have to be included in the analyzing process of healthcare system engineering problems. Reorganization of units in a healthcare system is one of the important problems in clinics. Most valuable units in hospitals do need reorganization or reinvestment to maintain their competitiveness. Reorganization has multi-dimensions; devices, informatics, and staff. Strategic reorganization of a hospital especially of its units is very important in the competitive environment owing to preserve the future cash flow of the foundation and increase in productivity. If the management invests in most valuable units, it provides improvement in not only cashflows but also in the quality perception.

Reorganization process of healthcare systems are rarely studied by the scientist in contrast to its significance. Paschoal and Castilho (2010) offered a structuring process for the Surgical Center including Materials Management System at the University Hospital of the University of Sao Paulo. Visualization and simulation tools utilization in reorganization and redevelopment of a hospital based on a spatial concept was examined by Persson et al. (2014). Bode (2015) studied the

reform processes of hospital care in Mexico and Germany. He searched paradoxical consequences of reorganization of healthcare. Leo et al. (2016) investigated the assessment of centrically allotment based on triage at the Department of Epidemiology of the Regional Health Service of Lazio, Italy. They claimed the offered method was much more efficient and effective.

In this section, explanatory definitions about unit investment planning in healthcare and ordering of these investment are provided in Sect. 17.5.1. Procedure of analytic network process (ANP) that is contained in operations research issue is given in Sect. 17.5.2. Section 17.5.3 implies a numerical application of the offered method in the selection process among units in a private hospital in Turkey.

17.5.1 MCDM in Unit Investment Ordering

Multiple-Criteria Decision Analysis (MCDA) or mostly used form; Multiple-Criteria Decision-Making (MCDM) is one of disciplines of operations research. One of the main criteria is frequently price or cost criterion. Qualitative criteria typically refers some measure of quality those are against cost criterion. Analytic hierarchy process (AHP), Analytic network process (ANP), Technique for the Order of Prioritization by Similarity to Ideal Solution (TOPSIS), ELECTRE (ELimination Et Choix Traduisant la REalite), Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE), VIKOR (VlseKriterijumska Optimizacija I Kompromisno Resenje), and Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH) are commonly used techniques in MCDA.

Multi-criteria decision making methods began finding fields at an increasing rate in healthcare systems based on multiple and conflicting criteria. Some of the MCDM utilization in healthcare are as follows: Lu et al. (2013) offered a hybrid MCDM method to assess the elements that effect the adoption of radio frequency identification (RFID) technology in Taiwan's healthcare system. Zeng et al. (2013) developed VIKOR method which exists among MCDM methods to assist the decision making procedure in the evaluation of Iodine deficiency disorder (IDD) mentoring efficiency.

Unit investment decisions are reorganization of a healthcare system and it is ranked among the strategic type of decisions. Like in many different sectors and problems, also in healthcare decision making problem decisions have vague and conflicting criteria. Sometimes those defined criteria may interrelated with each other. By those reasons, in this study a multi-criteria decision making method called ANP is offered as a solution procedure for unit reorganization investment decision problem.

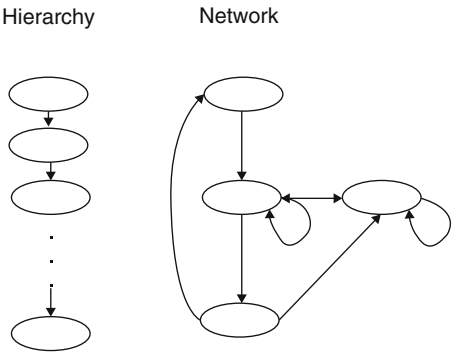

Fig. 17.8 Hierarchy and network illustration

17.5.2 ANP Method

Analytic network process (ANP) was built up by Saaty (1980, 2001) which can be utilized for analyzing network model impressions. ANP combines the assessments of all decision makers into an eventual decision making process by pair-wise comparisons of alternatives, without having to reveal their utility functions on subjective and objective criteria.

The general form of analytic network process said ANP is an extensive decision making method that deals with the consequence of interconnection and feedback among groups of elements. ANP method offers more sophisticated interrelationships among decision grades and features. Having no requirement of strictly hierarchical structure is the main advantage of ANP and that technique can proceed problems possessing sophisticated mutual relations among criteria. Therefore, ANP can tackle with the hardships of real-world problems better than AHP. Figure 17.8 characterizes a hierarchy and a network with mutual relations among criteria and feedback between determinants of a sophisticated framework.

ANP procedure consists of four principal steps which are structuring an issue into a fulfill set of network patterns, composing pair-wise comparisons to forecast precedences at each grade, constructing a supermatrix to indicate the influence precedence of elements and deciding on the established supermatrix. The principal aspect of ANP is that the supermatrix concept which reminds the Markov chain process. The difficulties in the interdependence characteristics of elements and constituents are coped by this supermatrix. The local priority vectors are interpolated to the convenient columns of the supermatrix to obtain global precedences of the system with interdependent impacts. A supermatrix is essentially a partitioned matrix, in which a link between the two nodes in a system represents each matrix section.

In this study four major steps of ANP method taken from Saaty (1996) main work are as follows:

Step 1. Constructing the model and problem configuration: The issue should be frankly defined and disintegrated into a conceivable arrangement e.g. a network.

Step 2. Matrices concerning pair-wise comparisons and precedence vectors: At each factor grade ANP decision elements are confronted with other same grade elements pair-wisely in respect of their control criteria, and the constituents themselves are also compared pair-wise according to their additive to the objective. The experts or decision makers are demanded to reply sequences of pair-wise comparisons in which two elements at once will be compared on account of what kind of contribution to their specific upper grade criterion. The notional significance levels are assigned on a scale of 1–9, where '1' indicates equal emphasis between the two elements and '9' displays extreme emphasis of one factor (row constituent in the matrix) against the other one (column constituent in the matrix). An opposing assess is appointed to the adverse comparison (i.e., $a_{ij} = 1/a_{ji}$) where a_{ij} represents the emphasis of the ith element compared to the jth element. Pair-wise comparison is generated in the structure of a matrix in ANP, and a local precedence vector can be acquired for predicting the relative significance associated with the components being compared by analyzing the following formulae:

$$A \cdot w = \lambda_{max} \cdot w \tag{17.30}$$

where the matrix of pair-wise comparison is indicated by A, eigenvector is displayed by w, and the largest eigenvalue of A is denoted by λ_{max}. If a consistency matrix is displayed by A, then X (i.e. eigenvector) can be specified using

$$(A - \lambda_{max} I)\, X = 0 \tag{17.31}$$

The consistency index (CI) and consistency ratio (CR) are offered by Saaty (1980) to confirm the consistency of the comparison matrix. These indexes are stated respectively as follows:

$$CI = (\lambda_{max} - n) \,/\, (n - 1) \tag{17.32}$$

$$CR = CI/RI \tag{17.33}$$

where *RI* indicates the average consistency index for innumerable coincidental entries of same-order opposite matrices. If $CR \leq 0.1$, then the forecast is approved; if not, a new comparison matrix is importuned until $CR \leq 0.1$.

Step 3. Formation of Supermatrix: The supermatrix notion is analogous to the Markov chain process as Saaty indicated. To provide global precedences in a system implicating interdependent effects, the local precedence vectors are enrolled into the convenient columns of a matrix, commonly a supermatrix. Herewith, a partitioned matrix is named as a supermatrix, where each matrix section denotes a relevance between two constituents in a system. Assume the constituents of a decision system be C_k; $k = 1, \ldots, n$, where each constituent k has m_k elements, depicted by $e_{k1}, e_{k2}, \ldots, e_{km_k}$. Then, group and locate the local precedence vectors derived in Step 2 in convenient positions in a supermatrix. That situation is grounded on the flow of effect from a constituent to itself or from one constituent to another.

A general form of a supermatrix can be arranged as follows:

$$W = \begin{array}{c} \\ C_1 \\ \\ \vdots \\ C_k \\ \\ \vdots \\ C_n \\ \\ \end{array} \begin{array}{c} \begin{array}{cccccccccc} C_1 & \cdots & C_k & \cdots & C_n \end{array} \\ \begin{array}{cccccccccc} e_{11} & \cdots & e_{m_1} & \cdots & e_{k1} & \cdots & e_{km_k} & \cdots & e_{n1} & \cdots & e_{nm} \end{array} \\ \begin{bmatrix} W_{11} & \cdots & W_{1k} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ W_{k1} & \cdots & W_{kk} & \cdots & W_{kn} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ W_{n1} & \cdots & W_{nk} & \cdots & W_{nn} \end{bmatrix} \end{array} \begin{array}{c} e_{11} \\ \vdots \\ e_{1m_1} \\ \vdots \\ e_{k1} \\ \vdots \\ e_{km_k} \\ \vdots \\ e_{n1} \\ \vdots \\ e_{nm_n} \end{array} \tag{17.34}$$

If the criteria are related to each other, W_{22} of W_n would represent the interdependency, and the supermatrix would be (Saaty 1996);

$$W_n = \begin{bmatrix} 0 & 0 & 0 \\ w_{21} & w_{21} & 0 \\ 0 & w_{32} & I \end{bmatrix} \tag{17.35}$$

Especially, if interrelatedness among the elements in a constituent or between two constituents occurs any zero in the supermatrix consisting of a switch by a matrix should be performed. At the end of a pair-wise comparison matrix of the row constituents with respect to the column constituent an eigenvector can be acquired. All the components in the first block of that column are multiplied by the initial input of the respective eigenvector, all the elements in the second block of that column are multiplied by the second input, and so on. Finally, each block consisting each column of the supermatrix is weighted, and the output is entitled the weighted supermatrix.

Step 4. The best of the alternatives selection: If the supermatrix effectuated in Step 3 contains the entire network, the precedence alternative weights can be detected in the alternatives column in the normalized supermatrix. Conversely, if a supermatrix is just occurred of interrelated constituents, further computations have to be made to acquire the overall precedences of the alternatives. The chosen alternative should be the one having the highest overall precedence.

17.5.3 An Example of Private Hospital in Turkey

As one of the vital sectors for the society, the health sector covers public and private hospitals. Since private hospitals are the companies with the aim to benefit, it is important to decide the investment decisions to maximize the profit of the company. The signifying difference between private hospitals and ordinary businesses is they occupied with human as the working constituent. Therefore, it is highly important to define the concept of quality in hospitals. So we can say that hospitals are businesses whose quality and regulations are vital while the ultimate goal is to profit. Especially in the last decade, the strategic decision of investment in hospitals, i.e., between the units of the hospital, take importance more than ever.

The hospital discussed in that section is a semi-private hospital which some special services are reimbursed by SGK (Social Security Administration in Turkey) however alternative units processed are operated fully private. This is important because we need all people may be at equal levels. The attractiveness of a field whose prices are reimbursed by SGK may be more than other areas. So it would not be a correct comparison.

The problem is that an investment will be made to one of the hospital units. It will be a rationalization of investment which means that the hospital does not want to establish new units. It just wants to enhance the money that developing a unit it already has. In deciding to invest in which unit we will determine the criteria affecting the hospital unit selection. The ANP method is used as the MCDM for the evaluation of hospital units.

17.5.3.1 The Identification of Criteria and Sub-criteria

The objective of our problem is prioritization of hospital units for the investment objective by using financial criteria and, in particular, non-financial. For this, we will first identify the qualitative and quantitative criteria that affect the process of selection with the help of Brain storming method, and of course other techniques like Delphi method could be used if need be. The units selected have the largest facilities, those are radiology, biochemistry, and orthopedics and traumatology. The criteria and sub-criteria are selected and eliminated according to the experience of the chosen hospital's top management. Criteria for investment decision are as follows:

Competitive advantage criterion consists of seven sub-criteria, development of the market criterion occurs two sub-criteria, technical background criterion consists of three sub-criteria, criterion management refers two sub-criteria, and flexibility criterion has two sub-criteria.

And finally income criterion which means profit for private hospital, of course it is based on the profit of the unit itself monthly. All these criteria, sub-criteria, and explanations of them are stated in Table 17.1.

Table 17.1 Criteria and explanations

Criteria & Sub-criteria	Explanation
Number of examinations	Average number of medical examinations made (those examinations should form at least 80% of all the facts examinations) for each patient in a month. It is not the number of patients because a patient can take several services in each unit. Number of examinations is related to amount of bed sub-criteria because longer the patient stays in the hospital, she/he will need more examination probably. Then, if the technical background of unit is strong, then more patients will prefer instead of other hospitals. In radiology unit, it is the number of radiography, magnetic resonance, angiography, tomography, and ultrasonography forming 85% of the examinations done in the radiology unit. In biochemistry, this is the number of routine examinations made by the unit of biochemistry. There are 16 examinations routine forming the 85% of examinations done in the field of biochemistry. In orthopedics and traumatology unit, it is the number of consultations and surgeries
Demand evaluation	The number of patients expected to come for next year. Demand evaluation is related to the ownership of new techniques or technologies, quality, price, length of service and technical background as these criteria increase the attractiveness of the unit
Ownership of new Techniques or technology	The distinction of the hospital from its rivals in sanitary technique and technological equipment
Quality	The quality of personnel service or technical equipment
Amount of bed	Number of beds in the hospital reserved for the mentioned units
Price	The average price of services provided for an individual. This is an important competitive advantage because if prices are low enough with the same quality of rivals, so it is a good reason to choose this hospital
Length of service	The average duration of the services provided by the unit
Trend for R&D studies	Every medical unit has R&D studies worldwide and researches slightly increase the probability of discovery of a new technique or technology that can create increased demand
Market potential	Position of the hospital in the market
Know-how	This is an important criterion since it has a great effect on the reputation of doctors and then the hospital
Team squad	The adequate number of doctors, their capability and coherence of the team. In fact, these criteria indicate the quality of the team

(continued)

Table 17.1 (continued)

Criteria & Sub-criteria	Explanation
Technical equipment	That is the machines and equipment used by the units. In radiology, it refers the machines. In biochemistry, this means the equipment utilized in laboratories. In both units, this is an important criterion because their work depends on the technical equipment more than orthopedic and trauma work
Number of alliance	This is the contract between different hospitals to change the patient because every hospital either has lack of technical equipment, lack of capacity or sometimes lack of know-how. This exchange of patient brings them the patient satisfaction since they inform patients instead of rejecting
Possibility of alliance	It is the possibility of making alliance between hospitals. This criterion depends on the vision of the hospital's top managers
Unit reaction under the emergency situation	The maximum capacity and the unit reaction to emergency situation
Maximum resources	Ability to serve simultaneously. This criterion depends on amount of beds, number of doctors and number of machines
Income	Profit for private hospital, of course it is based on the profit of the unit itself monthly

As taken from the hospital; number of examinations are 17,000; 150,000; 900 respectively for units radiology, biochemistry, and orthopedics and traumatology. Then only unit orthopedics and traumatology has 15 beds. The prices per patient are ৳1,200; ৳500; ৳4,500 for units radiology, biochemistry, and orthopedics and traumatology, respectively. Monthly profits for year 2012 provided by the hospital for units radiology, biochemistry, and orthopedics and traumatology were ৳350,000; ৳1,350,000, ৳1,400,000, in order. We normalize the quantitative data and fill the survey with the normalized data. So, the quantitative data does not depend on the opinion of experts. These numbers are the real data that was taken from the hospital we studied. The only financial criterion is profit because they had data rows to other financial calculations but there was no such database in the hospital.

The network of sub-criteria depicts the connections between the sub-criteria, the connections automatically eventuate between criteria. Then there is compulsory link between all sub-criteria and alternatives and that all criteria have been committed to this objective. Relations among criteria were asked to the experts with the help of survey utilization and they are presented at the third column of Table 17.2. When we enter all the relationships data to the Super Decision program, the network is obtained as in Fig. 17.9.

Table 17.2 Criteria and relations among them

Criteria	Sub-criteria	Relation among criteria and sub-criteria
(C_1) Competitive advantage	(C_{11}) Number of examinations	C_{13}, C_{31}, C_{32}, C_{33}
	(C_{12}) Demand evaluation	C_{14}, C_{15}, C_{16}, C_{17}, C_{31}, C_{32}, C_{33}
	(C_{13}) Amount of bed	C_{11}, C_{51}, C_{52}
	(C_{14}) The ownership of new techniques or new technology	C_{11}, C_{21}, C_{31}
	(C_{15}) Quality	C_{12}, C_{16}
	(C_{16}) Price	C_{12}, C_{15}, C_{17}
	(C_{17}) Length of service	C_{31}, C_{32}, C_{33}
(C_2) Development of the market	(C_{21}) Trend for R&D Studies	C_{12}, C_{22}
	(C_{22}) Market potential	C_{14}, C_{21}
(C_3) Technical background	(C_{31}) Know-how	C_{41}, C_{61}
	(C_{32}) Team squad	C_{51}, C_{52}, C_{61}
	(C_{33}) Technical equipment	C_{61}
(C_4) Management	(C_{41}) Number of alliance	C_{11}, C_{31}, C_{32}, C_{33}
	(C_{42}) possibility of alliance	C_{12}, C_{22}, C_{31}, C_{32}, C_{33}
(C_5) Flexibility	(C_{51}) Unit reaction under the emergency situation	C_{52}
	(C_{52}) Maximum resources	C_{51}
(C_6) Income	(C_{61}) Profit monthly	C_{11}, C_{15}, C_{16}, C_{32}, C_{33}, C_{41}

17.5.3.2 Evaluation of the Problem by ANP

A private hospital (established in 1997) that is a part of a large hospital chain declared that it had cash surplus in 2012, however they were arguing reorganization. The hospital has 50,254 m^2 closed area, 219 beds for inpatients and advanced oncology center. A collaboration was made with the management, and then a strategic reorganization plan is offered with the help of MCDM method called ANP. A pareto analysis was made among units and it stated that these three units; radiology, biochemistry, and orthopedics and traumatology bring 80% of the profit. Therefore, a selection was to be made among those three alternatives to invest for strategic reorganization.

Radiology denotes the set of diagnostic and therapeutic methods using X-Rays, or more generally using radiation. Radiology as a medical specialty, concerning the following areas: conventional radiology, mammography, computed tomography (CT), interventional radiology, magnetic resonance imaging (MRI), and ultrasound. Radiology is a unit that serves all departments of hospital. Even if patients do not come to the hospital for radiological services, there is a great chance that the doctors they will, demand for radiological examinations. The large number of patients per month, the taking of any hospital patient for radiological examinations, and the need of expensive machines are the other characteristics of radiology unit that we have chosen to evaluate.

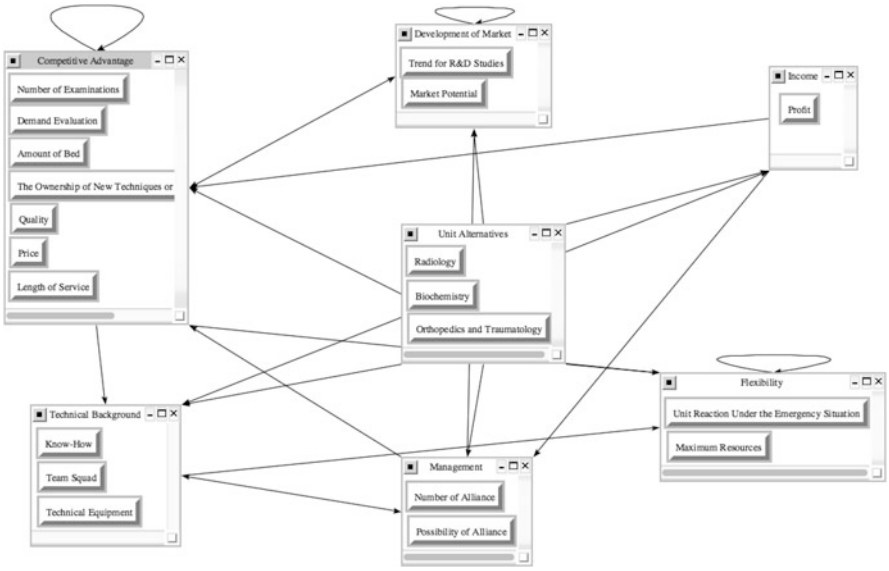

Fig. 17.9 Criteria network

Biochemistry is the scientific discipline that studied the chemical reactions taking place within the living and therefore within cells. The hospitals have laboratories to perform the analyzes they need to function. These analyzes are performed using laboratory techniques. All biochemistry services are given in laboratories. Then, one can simply say that the biochemistry costs will be lower than other units. The other characteristics for which biochemistry had been chosen as an alternative are large number of patients per month, the attraction of high patient throughout the hospital for biochemical examinations (actually more frequently than radiology), biochemical tests are the most routine examinations, and low costs, so the profit is high.

Orthopedics is a surgical specialty performed by an orthopedic surgeon. It concerns the treatment of all diseases of the musculoskeletal system (bones, joints, muscles, tendons and nerves). It includes the surgical treatment of upper limb disorders (shoulder, elbow and hand), the lower limbs (hip, knee and foot), and the spine. The other characteristics of orthopedics and traumatology unit referred as an alternative are at first, the reputation of orthopedics in the hospital chosen is high thanks to orthopedic doctors and the second; the astronomically high prices.

When we determine all the criteria, sub-criteria, alternatives and relations between them, Super Decision software gives us questions to assess all the criteria, sub-criteria and alternatives. Hospital top managers fulfill the surveys according to their expert advice. We gave a short training of the theoretical base of the ANP method. Surveys appears as in Fig. 17.10. The sample of the survey seen in the example is for the criterion of “Competitive Advantage”. The managers compare the other criteria for the criterion of “Competitive Advantage”.

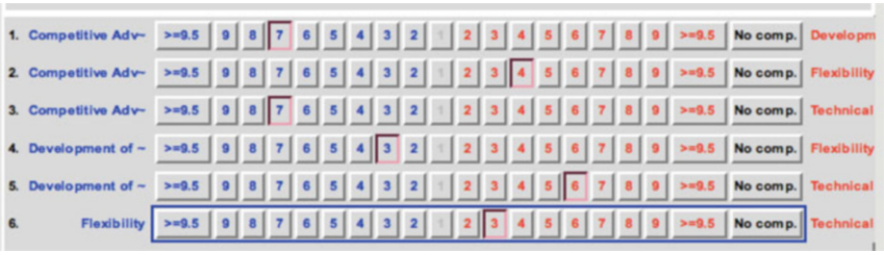

Fig. 17.10 Sample survey done by hospital managers

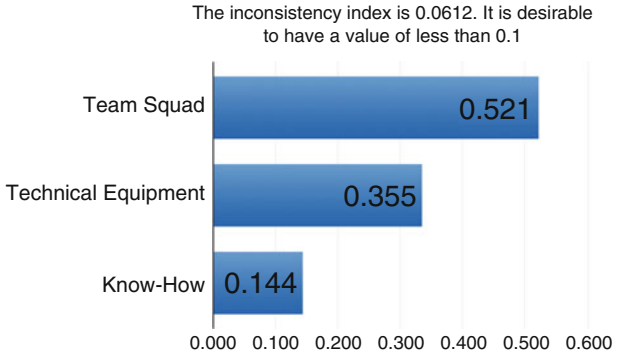

Fig. 17.11 CI example

After filling out the surveys, we have to control 'consistency index' which must be smaller than 0.1. This is necessary because the surveys are filled by human which means there may be still no consistent respond. In Fig. 17.11, it can be easily seen an example of the consistency index and it is 0.0612 for instance. Consistency indicators are controlled and verified that they have values less than 0.1. Results of these comparisons develop supermatrix, however because of page limitations we could not present whole matrix.

We transform the supermatrix into a weighted supermatrix where the weight of each column is equal to one. Then, weighted matrix is raised to the power unknown until the limit matrix is obtained. The priorities of alternatives can be seen from the matrix where each column is equal. According to the limit matrix two solutions are obtained: at first it can be said which alternative is the best to reorganize/invest. Second, it can be deduced the most important sub-criterion from all the others. Figure 17.12 shows the result obtained by the limit matrix for alternatives and gives percentages for each alternative. Then, we can say that it is best to reorganize or invest in orthopedics and trauma unit.

At the end of the application there have been two results. The first one is at the criteria level. According to the relationships those were composed and in accordance with the results of surveys completed by executives, we obtained the criterion 'profit' as the most important among the criteria as can be seen from Fig. 17.13.

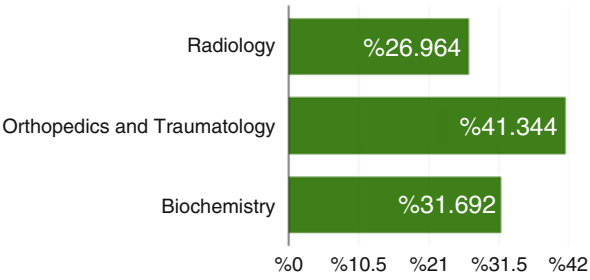

Fig. 17.12 Priorities of alternatives

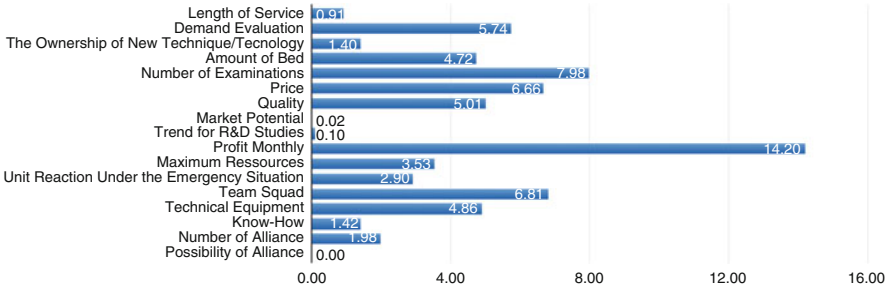

Fig. 17.13 Priorities of criteria and percentages

This was an expected outcome since the first goal of almost every private business is to make profit though it is a public area. If one puts the results in a descending order according to their weights; the following criteria are obtained as the top eight most important ones: Profit, Number of Examinations, Team Squad, Price, Demand Evaluation, Quality, Technical Equipment, Amount of Bed.

Among these eight criteria, the first four were the expected results but what is surprising is the fact that the quality is ranked as the sixth important criterion. Whilst the concept of quality is very important for hospitals. So we can say that all hospitals in the market are trying to have a high quality and it is not the criterion that creates the difference among the competitors. Since quality has become a minimum criterion to have, it is not the criterion that will create competitive advantage in the market. At the alternative level, as a first choice it is better to make reorganization in orthopedics and trauma, the second one is the biochemistry and the last one is in the radiology unit. Before making this prioritization, it was known that the biochemical profit and orthopedics were too close. Before this analysis, we could have doubts about which unit to reorganize. But after making prioritization we saw that orthopedics and traumatology unit is better to organize.

17.6 Medical Decision Making: Device Selection Problem

In healthcare system offering health services is performed by organizing institutions, people, and resources to fulfill the health requirements of target populations. In limited sense healthcare system as utilized in that chapter, refers individual health care services for people and their families, available at district centers, clinics, hospitals, and similar organizations, in doctors' offices, and in patients' own houses. The health service includes protection of health, prevention of diseases, diagnosis and treatment of disease, and rehabilitation. Even though all these components are important for health care, diagnosis is critical for right and on time treatment. In recent times, medical imaging have become crucial element during the diagnosis process. As time goes by medical imaging devices began to continuous improvement. However due to financial resource constraints, acquiring whole updated imaging devices may enforce the establishment of any type of healthcare organization. A ranking of medical imaging devices or selection of the one among them will recover financial plans of the institution.

Most of the medical imaging devices operate with the help of the computers and also there is information process between these devices as a whole and physicians during the diagnosis and treatment process. Informatics involves the interaction between humans and information besides the building of the interfaces among organization, technology and system. Venot et al. (2014) described the medical informatics as medical knowledge modeling and reasoning, covering exchange of paper medical records with electronic records and supporting or substituting doctors in diagnostic or treatment procedures. In the ordering or selection process of medical imaging devices the procedure includes medical informatics.

The preference of medical imaging devices is a compelling problem in case of medical decision making and budget planning. The doctors request new generation devices to exert right and reliable diagnosis. Despite that the contributories endeavor to invest in medical devices by their point of view or their priorities that is liable to both the hospital requirements and specifications. Both financial and technological aspects have to be appraised with several features in the selection of a medical device.

Economic aspect of medical device selection (MDS) problems are dealt in several studies. Ekwueme et al. (2002) investigated seven different injection devices for their economic costs and their jeopardy of iatrogenic transmission in sub-Saharan Africa. Vallejo-Torres et al. (2008) offered an adoption of the Bayesian method for their implementation of directly affording investment decisions in a commercial regulation from ancient phases of the improvement of novel medical devices. Sloan (2010) indicated a Markov decision process structure in the utilization of whether to use new or reprocessed devices decision. Rehana et al. (2013) denoted a prospective study on medical imaging devices for a six month period in computation of unit cost of radiological investigations. Gilard et al. (2013) represented a methodology of scholarly assessment of medical devices. They incorporated procedures in the French National Health Insurance system with their view of pricing and financing. Agapova et al. (2014) provided a benefit-risk assessment (BRA) for pharmaceuticals or medical devices, especially for imaging technologies.

The medical imaging devices effective for doctors and associated with the hospital investors and hospital engineers are focused in this study. We proposed an approximation which supports one of the multi-criteria decision making methods VIKOR to assess the selection of these devices.

The organization of this section is as follows: In the following subsection multi-criteria decision making methods are explained and especially offered one VIKOR is presented. The second subsection provides the reason of why to use MCDM in medical decision making analysis. The third subsection consists of a real case example for a private hospital in Turkey for the selection of medical imaging devices.

17.6.1 MCDM in Medical Decision Making

An explanation about MCDM and various methods used in multi-criteria decision analysis were discussed in Sect. 17.5.1. Since the MCDM methods are effective and scientific techniques they are mostly preferred in all field of healthcare. Every method has its own advantage and disadvantage. According to the characteristic of the problem the most appropriate one would be chosen in order to get best result. Selecting the suitable method is ensured by not only listing the pros and cons of the related methods but also trusting the researcher's experience.

In hospital engineering, device choice process is one of the hardest problems that could be analyzed by MCDM. A proper distribution of investment among devices by the courtesy of budget planning and priorities provides flexibility. Taghipour et al. (2011) investigated an MCDM to configure precedence in medical devices according to their critical order. Cho and Kim (2003) introduced the utilization of analytic hierarchy process (AHP) in evaluating chosen medical devices and materials for donations by the Korean Ministry of Health and Welfare. Pecchia et al. (2013) submitted an AHP analysis for a novel Computed Tomography (CT) scanner in which they offered the hierarchy of 12 needs and gathered 4 homogenous bunches. Ivlev et al. (2014) offered a classification of MCDM methods for selecting medical equipment and they tried to identify the most convenient one for performing management tasks in limited resource environment.

17.6.2 VIKOR Method

The VIKOR method is situated amongst multi-criteria decision making methods. Opricovic and Tzeng (2004) offered a technique for analyzing decision issues with conflicting and incommensurate criteria. They supposed that compromise solution is admissible for conflict resolution which means a solution that is the closest to the ideal is asked by the decision maker. All structured criteria are assessed by the decision maker for the evaluation of alternatives. The essential benefit of VIKOR is investigating the closest to the ideal solution (i.e. compromise solution) and ordering alternatives from best to worse.

VIKOR method propounds differently from other methods, aggregation functions, normalization method, and compromise solution all together. Calculating the optimal node in the VIKOR is based on the certain measure of "closeness" to the positive ideal solution. For this reason, in case of when the decision maker (DM) desires maximum profit but simultaneously decision risk is less important for the DM this method is the most appropriate one.

As distinct from other MCDMs, VIKOR uses linear normalization and puts forth an aggregating function that indicates distance from the ideal solution. This function refers clustering of criteria, notional emphasis of criteria, and a stabilization among individual and total contentedness. Though the essential concern is the reference point for decision making process, being much possible close to the ideal is sensible explanation of choice of human. The closest point to the ideal solution signifies the alternative with highest rank in VIKOR technique. Besides ordering, that technique suggests a compromise solution with an advantage rate.

The algorithm of the VIKOR method is as follows:

The different J alternatives are depicted as $a_1; a_2; \ldots; a_J$. For alternative a_j, the rating of the ith aspect is defined by f_{ij}, i.e. f_{ij} is the assess of ith criterion function for the alternative a_j; and number of criteria is depicted by n.

The important aspect of VIKOR technique is to compute the following form of $L_p - metric$:

$$L_{p,j} = \left\{ \sum_{i=1}^{n} [w_i (f_i^* - f_{ij}) / (f_i^* - f_i^-)]^p \right\}^{1/p}, 1 \le p \le \infty; j = 1, 2, \ldots, J. \quad (17.36)$$

Within the VIKOR method, to arrange the ranking measure $L_{1,j}$ (as S_j in Eq. (17.37)) and $L_{\infty,j}$ (as R_j in Eq. (17.38)) have to be calculated. The solution of $min_j\, S_j$ refers a maximum group utility ("majority" rule), and the solution of $min_j R_j$ refers a minimum individual regret of the "opponent". The feasible solution of the "closest" to the ideal F^* contains the compromise solution F^c, and compromise means a settlement built by mutual privileges. The compromise ranking algorithm VIKOR has the following steps:

(a) Determine the all criterion functions that is represented by the best f_i^* and the worst f_i^- values, $i = 1, 2, \ldots, n$. If a criterion is a benefit one then the ith function is demonstrated as:

$$f_i^* = \max_j f_{ij}, \quad f_i^- = \min_j f_{ij}$$

(b) Calculate the values S_j and $R_j, j = 1, 2, \ldots, J$, by the relations

$$S_j = \sum_{i=1}^{n} w_i (f_i^* - f_{ij}) / (f_i^* - f_i^-), \quad (17.37)$$

$$R_j = \max_j [w_i (f_i^* - f_{ij}) / (f_i^* - f_i^-)], \quad (17.38)$$

where the weights of criteria are denoted by w_i that expresses their notional significance.

(c) Calculate the Q_j values, where $j = 1, 2, \ldots, J$, by the relation

$$Q_j = v\left(S_j - S^*\right) / \left(S^- - S^*\right) + (1 - v)\left(R_j - R^*\right) / \left(R^- - R^*\right) \qquad (17.39)$$

where
$S^* = \min_j S_j, \quad S^- = \max_j S_j$
$R^* = \min_j R_j, \quad R^- = \max_j R_j$
and v is presented as the maximum group utility which can be referred as the weight of the strategy of "the majority of criteria", here $v = 0.5$.

(d) Sort the alternatives in decreasing order by ranking the values of S_j, R_j, and Q_j. These should be three ranking lists.

(e) Prefer the alternative $(a^{'})$ as a compromise solution that is sorted by the best measure of Q (minimum) if the following two provisions are ensured:

C1. "Acceptable advantage":

$$Q(a^{''}) - Q(a^{'}) \geqslant DQ \qquad (17.40)$$

where the alternative with second best position in the ranking list by Q is indicated as $(a^{''})$; $DQ = 1/(J-1)$; and the number of alternatives is represented by J.

C2. "Acceptable stability in decision making": Alternative $(a^{'})$ should also be included in the set of the best ranked by S_j or/and R_j. Inside of the decision making process, that offered compromise solution is consistent. The solution could be represented as: "voting by majority rule" (when $v > 0.5$ is needed), or "by consensus" $v \approx 0.5$, or "with veto" ($v < 0.5$). At that point, the maximum group utility (i.e. the weight of the decision making strategy "the majority of criteria") is displayed by v.

If one of the circumstances is not ensured, then compromise solutions cluster is offered, and this cluster should contain the following situations:

- If the only circumstance **C2** is not ensured, choose alternatives $a^{'}$ and $a^{''}$, or
- If circumstance **C1** is not ensured, alternatives $a^{'}, a^{''}, \ldots, a^{(M)}$; and $a^{(M)}$ is specified by the statement $Q\left(a^{(M)}\right) - Q\left(a^{'}\right) < DQ$ for maximum M (the situations of those alternatives are "in closeness").

The minimum value of Q represents the best alternative. However, the essential ordering outcome should include the compromise ranking list of alternatives, and "advantage rate".

17.6.3 A Case Study

Main healthcare services are provided by two major basic structures: public and private. In that case we offer a selection process of medical imaging devices for a private hospital. Apart from few examples, most of the private hospitals in Turkey are small or medium sized enterprises. For that kind of organizations every type resources are much more important according to large scale hospitals. It was contacted by a financier who wants to invest in medium sized hospital. In the initial phase the investor do not wish to expend on medical devices and also on informatics structure. However, some devices must be purchased in order to operate the hospital properly. That study seeks an answer to that problem. A committee concerning two medical specialists on imaging service, two radiologists, and a healthcare engineer has been created for the choice of medical imaging devices. Criteria selection, evaluation, and comments are the product of collaborative effort.

17.6.3.1 Criteria for Medical Imaging Device Selection and Alternatives

In hospital engineering, criteria selection phase is the critical stage of medical device selection process though it has significant effects on solution. Though diagnostic facilities and medical imaging devices have severe cost burden to attain health standards and fulfill trustworthy decision making, this expensive investment analysis has to be done scientifically.

After few meetings, a compromise decision about criteria has been made, and the criteria determined are depicted in Table 17.3. As can be seen from the table, there are 5 main clusters including hardware, software and imaging specifications separately, recording and archiving, and finally initial costs. With the main criteria of initial cost and fourteen sub-criteria the decision makers should evaluate fifteen criteria relatively to alternatives. Also they determined the weights of criteria according to their experiences.

In this section six different medical imaging devices are identified by the committee and us. We have ordered all medical imaging devices that should be in a full-fledged hospital, then we made a Pareto analysis which we seek devices with 20% meet 80% of imaging operations with regarding to historic data studied at the previous section. At the end of this analysis we have found alternatives for medical imaging device selection process as follows: Ultrasonography, X-Ray, Computerized Tomography (CT), Magnetic Resonance (MR), Scintigraphy, Positron Emission Tomography (PET).

17.6.3.2 Evaluation of MDS Problem by VIKOR

Every experts are consulted with survey about the evaluation of criteria taking the alternatives into consideration. At first, weighting assignment is requested; then the

Table 17.3 Criteria for medical imaging device selection

Criteria	Sub-criteria
(C_1) Hardware specifications	(C_{11}) Resolution
	(C_{12}) Physical properties (size-weight), portability
	(C_{13}) Power standards
	(C_{14}) Patient conformability
	(C_{15}) Acquisition properties
(C_2) Imaging specifications	(C_{21}) Monitoring
	(C_{22}) Digital standards
	(C_{23}) Integrability
(C_3) Recording and archiving	(C_{31}) Support for digital and analog radiology
	(C_{32}) Digital radiology features
	(C_{33}) PACS-Telemedicine
(C_4) Software Specifications	(C_{41}) Signal/Noise ratio
	(C_{42}) Multidimensional view
	(C_{43}) Artificial enhancements-augmented reality
(C_5) Initial cost of alternatives	

Table 17.4 Weight and criteria opinion of one of the experts

	w_i	Ultrasonography	X-ray	Tomography	MRI	Scintigraphy	PET
Resolution	16	5	10	30	30	10	15
Physical properties	15	30	10	20	20	10	10
Power standards	5	10	10	20	20	20	20
Patient comfort	1	35	20	15	10	10	10
Acquisition properties	10	25	15	20	20	10	10
Monitoring	5	25	15	15	20	10	15
Digital standards	1	15	15	25	25	10	10
Integrability	10	25	10	20	30	10	5
Support for A&D radiology	5	25	15	15	25	10	10
Digital radiology properties	5	20	15	20	20	10	15
PACS-Telemedicine	15	20	20	20	30	5	5
SNR ratio	1	15	15	20	25	15	10
Multidimensionality	5	15	10	20	30	15	10
Augmented reality-special features	1	10	10	30	30	10	10

criteria are asked. The economic data are gathered from a sales expert for an agency that vends imaging devices for hospitals, separately from those experts mentioned before. You can see the evaluation of one of the experts in Table 17.4. Due to confidentiality agreements, only one judgment result can be shared.

Table 17.5 Weights of criteria, average of expert opinions, and economic data

	w_i	Ultrasonography	X-ray	Tomography	MRI	Scintigraphy	PET
Resolution	12.2	12	13	26	26	11	12
Physical properties	10	28	19	13	14	14	12
Power standards	6	8	14	20	23	17	18
Patient comfort	3.4	28	24	15	11	12	10
Acquisition properties	11	17	14	24	25	9	11
Monitoring	5	24	13	20	23	9	11
Digital standards	4.4	14	18	23	26	9	10
Integrability	6.4	29	15	18	21	9	8
Support for A&D radiology	6	23	14	22	24	8	9
Digital radiology properties	7	17	19	22	21	9	11
PACS-Telemedicine	10	20	16	21	24	9	10
SNR	3.4	15	14	23	25	11	12
Multidimensionality	5.8	18	9	26	29	9	9
Augmented reality-special features	2.4	15	14	25	28	9	9
Initial costs (in thousand $)	7	60	175	700	1200	1390	850

Table 17.6 Calculated variables of S_j, R_j, and Q_j

	S_j	R_j	Q_j
Ultrasonography	0.4599	0.1139	0.5908
X-ray	0.6545	0.1057	0.6097
Tomography	0.2729	0.0937	0.1680
MRI	0.1627	0.0875	0
Scintigraphy	0.8746	0.1220	1
PET	0.8637	0.1139	0.8744

After consolidation of all the experts' judgements and economic data Table 17.5 is obtained as an initial table for VIKOR technique.

At the first step we have to compute $L_{p,j}$, with its sub-components $L_{1,j}$, and $L_{\infty,j}$ also known as relations; S_j and R_j, respectively. Equations (17.37) and (17.38) should be utilized for the calculation of cited values consecutively. For the selection or ranking process Q_j values have to be determined by Eq. (17.39) according to data in Table 17.5. Results of these calculations and sorting by the values of S_j, R_j, and Q_j in decreasing order are depicted in Tables 17.6 and 17.7, sequentially. Weight of the strategy of majority of criteria or better expressed as maximum group utility value "v" is accepted as 0.5.

$Q(minimum)$ depicted as $Q(a')$ equals zero as implies from Table 17.6. Alternative with the second position in ranking list denoted as $Q(a'')$ equals 0.1649 which is lower than DQ, so condition 1 is not satisfied. Then we should investigate the second condition shown in step (e). When we check $Q(a')$ values from Table 17.7

Table 17.7 Sorting of alternatives according to S_j, R_j, and Q_j values

S_j	R_j	Q_j
Scintigraphy	Scintigraphy	Scintigraphy
PET	Ultrasonography	PET
X-ray	PET	X-ray
Ultrasonography	X-ray	Ultrasonography
Tomography	Tomography	Tomography
MRI	MRI	MRI

one can see alternative ($a^{'}$) is the best ranked by S_j and R_j also. The second condition is provided. MRI must be selected as a best alternative.

Though there is a conflict between conditions we have to find a compromise solution. We have to calculate $a^{'}, a^{''}, \ldots, a^{(M)}$ values to reach the accurate solution. When the computation is done one can see only the second alternative $a^{''}$ is below DQ, so we can deduce that alternative tomography should be included in the compromise solution.

The result really makes sense; though the budget is restricted for the investment the experts are thinking the best alternatives for whole diagnosis processes. The selected alternatives MRI and Tomography could be applied to every part of the body.

17.7 Conclusion

Soft computing tools based on optimization techniques have become pervasive in medicine, life sciences and pharmaceutics. Recently, algorithms and decision making tools resolved many challenging issues in healthcare especially in medical imaging, bioinformatics and hospital engineering. Nevertheless, the chaotic nature of biological systems and the inconsistency of theoretical models in the application area create an important feedback for further developments.

In this chapter, we intended to create a general point of view about healthcare problems from micro level; bioinformatics and image analysis to macro level; hospital organization and medical decision making. A brief review of medical informatics was presented in the introduction regarding the whole aspects of the entire topic. For this purpose, it is emphasized that a multidisciplinary point of view would be given by different approaches in applied problems of medicine.

In micro level, an essential problem of virology, pandemic analysis was studied through sequence alignment. H1N1 influenza is a severe disease which may effect entire world with temporal cycles regarding to quick and unpredicted evolution. The similarity measure in a virus collection is important for drug development and delivery systems. In this section, we studied this problem by two different approaches; global; NM and local; SW algorithms. We noted that even in a limited

region a global technique may not be sufficient to cover the entire similarity value due to quick evolution of influenza.

Afterwards, two different problems of medical image processing were presented using different optimization tools. Even though an automatic segmentation may not be satisfactory due to several conditions (noise, acquisition, protocol etc..) clinicians prefer to apply an algorithmic approach to compare their findings before taking a decision. In clinical analysis, variational methods were found satisfactory in order to evaluate medical images. In this part, we applied this technique; snakes on skull segmentation for multimodal purposed. In general, image registration problem was found indispensable where patient motion or image deformation were higher. Even if, modern approaches might decrease these artifacts using statistical approaches, radiologists apply a preprocessing step to match source and target images. This process was studied using ICP algorithm based on Levenberg-Marquart optimization. The multimodal registration was performed using already segmented slices of the skull.

Furthermore, healthcare information systems were studied through a real market case of hospital engineering. A strategic reorganization among hospital units is one of the most important problems in medical engineering. Radiology, orthopedics and traumatology, and biochemistry units were selected among whole units of hospital because of these units were producing 80 percent of work in that healthcare system. A reorganization investment thought by the administration of the hospital was evaluated with an MCDM approach. And then we obtained a solution of funding the orthopedics and traumatology unit. With the help of ANP method staying between close options was denied and attained reasonable solution against heuristic decisions referenced in business life.

Then, medical decision making was incorporated to the optimal selection of medical imaging facilities. Medical device selection topics became popular in the last decade according to its huge place in hospital investment. Imaging devices improve the quality of diagnosis process though it is vital for the whole quality of medical system. We remark that popular optimization routines were adapted to these problems. An initial cost integrated VIKOR approach was performed to the medical device selection problem. With the nature of cited approach two alternatives called MRI and Tomography were selected against other methods in the literature which was more logical in said problem.

Finally, the growth of multidisciplinary effects in medicine allows us to apply other engineering tools; decision making, machine learning and mathematical background; different spaces and metrics with the same purposes in future medical applications. Medical big data problems, nanomedicine, pharmaceutical studies, e-health applications, cellular imaging, medical robotics and instrumentation are other important fields which might be included in the future studies about medical informatics.

References

Agapova M, Devine EB, Bresnahan BW, Higashi MK, Garrison LP (2014) Applying quantitative benefit–risk analysis to aid regulatory decision making in diagnostic imaging: methods, challenges, and opportunities. Acad Radiol 21(9):1138–1143

Besl PJ (1992) McKay ND Method for registration of 3-d shapes. In: Robotics—DL tentative. International Society for Optics and Photonics, pp 586–606.

Bode I (2015) A 'world culture'of institutional ambiguity? Comparing the reorganization of hospital care in Germany and Mexico. Curr Sociol 63(3):411–431

Castellani U, Bartoli A (2012) 3d shape registration. In: 3D imaging, analysis and applications. Springer, Berlin, pp 221–264

Cho K-T, Kim S-M (2003) Selecting medical devices and materials for development in Korea: the analytic hierarchy process approach. Int J Health Plan Manag 18(2):161–174

Duan F, Yang Y, Li Y, Tian Y, Lu K, Wu Z, Zhou M (2014) Skull identification via correlation measure between skull and face shape. IEEE Trans Inform. Forensics Secur 9(8):1322–1332

Ekwueme DU, Weniger BG, Chen RT (2002) Model-based estimates of risks of disease transmission and economic costs of seven injection devices in sub-Saharan Africa. Bull World Health Organ 80(11):859–870

Fitzgibbon AW (2003) Robust registration of 2d and 3d point sets. Image Vis Comput 21(13):1145–1153

Fulekar M (2009) Bioinformatics: applications in life and environmental sciences. Springer Science & Business Media, Berlin

Gallaher WR (2009) Towards a sane and rational approach to management of influenza h1n1 2009. Virol J 6(1):1

Gilard M, Debroucker F, Dubray C, Allioux Y, Aper E, Barat-Leonhardt V, Brami M, Carbonneil C, Chartier-Kastler E, Coqueblin C et al (2013) Scientific evaluation and pricing of medical devices and associated procedures in France. Thérapie 68(4):201–208

Gooya A, Liao H, Matsumiya K, Masamune K, Masutani Y, Dohi T (2008) A variational method for geometric regularization of vascular segmentation in medical images. IEEE Trans Image Process 17(8):1295–1312

Handl J, Kell DB, Knowles J (2007) Multiobjective optimization in bioinformatics and computational biology. Trans Comput Biol Bioinform 4(2):279–292

Ho S, Bullitt E, Gerig G (2002) Level-set evolution with region competition: automatic 3-d segmentation of brain tumors. In: 16th international conference on pattern recognition, 2002. Proceedings, vol 1. IEEE, New York, pp 532–535

Ivlev I, Kneppo P, Bartak M (2014) Multicriteria decision analysis: a multifaceted approach to medical equipment management. Technol Econ Devel Econ 20:576–589

Kass M, Witkin A, Terzopoulos D (1988) Snakes: active contour models. Int J Comput Vis 1(4):321–331

Koppstein D, Ashour J, Bartel DP (2015) Sequencing the cap-snatching repertoire of h1n1 influenza provides insight into the mechanism of viral transcription initiation. Nucleic Acids Res 43(10):5052–5064

Leo G, Lodi A, Tubertini P, Di Martino M (2016) Emergency department management in Lazio, Italy. Omega 58:128–138

Li C, Huang R, Ding Z, Gatenby, JC, Metaxas, DN, Gore JC (2011) A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. IEEE Trans Image Process 20(7):2007–2016

Lu M-T, Lin S-W, Tzeng G-H (2013) Improving RFID adoption in Taiwan's healthcare industry based on a dematel technique with a hybrid MCDM model. Decis Support Syst 56:259–269

Maurer CR, Aboutanos GB, Dawant BM, Maciunas RJ, Fitzpatrick JM (1996) Registration of 3-d images using weighted geometrical features. IEEE Trans Med Imaging 15(6):836–849

Opricovic S, Tzeng G-H (2004) Compromise solution by MCDM methods: a comparative analysis of vikor and topsis. Eur J Oper Res 156(2):445–455

Paschoal MLH, Castilho V (2010) Implementation of computerized material management system at the University of São Paulo university hospital. Revista da Escola de Enfermagem da USP 44(4):984–988

Pecchia L, Martin JL, Ragozzino A, Vanzanella C, Scognamiglio A, Mirarchi L, Morgan SP (2013) User needs elicitation via analytic hierarchy process. A case study on a computed tomography scanner. Eur J Oper Res 13:2

Persson J, Dalholm EH, Johansson G (2014) Informing hospital change processes through visualization and simulation: a case study at a children's emergency clinic. Health Environ Res Design J 8(1):45–66

Rehana K, Tabish SA, Gojwari T, Ahmad R, Abdul H (2013) Unit cost of ct scan and MRI at a large tertiary care teaching hospital in North India. Health 5(12):2059

Saaty TL (1980) The analytic hierarchy process: planning, priority setting, resources allocation. McGraw Hill, New York

Saaty TL (1996) Decision making with dependence and feedback: the analytic network process, vol. 4922. RWS publications, Pittsburgh

Saaty TL (2001) Analytic network process. In: Encyclopedia of operations research and management science. Springer, Berlin, pp 28–35

Sloan T (2010) First, do no harm? A framework for evaluating new versus reprocessed medical devices. J Oper Res Soc 61(2):191–201

Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL, Raghwani J, Bhatt S et al (2009) Origins and evolutionary genomics of the 2009 swine-origin h1n1 influenza a epidemic. Nature 459(7250):1122–1125

Su L-M, Vagvolgyi BP, Agarwal R, Reiley CE, Taylor RH, Hager GD (2009) Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3d-ct to stereoscopic video registration. Urology 73(4):896–900

Taghipour S, Banjevic D, Jardine AK (2011) Prioritization of medical equipment for maintenance decisions. J Oper Res Soc 62(9):1666–1687

Tsai A, Yezzi A Jr, Willsky AS (2001) Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification. IEEE Trans Image Process 10(8):1169–1186

Vallejo-Torres L, Steuten LM, Buxton MJ, Girling AJ, Lilford RJ, Young T (2008) Integrating health economics modeling in the product development cycle of medical devices: a Bayesian approach. Int J Technol Assess Health Care 24(04):459–464

Venot A, Burgun A, Després S, Degoulet P (2014) Medical informatics as a scientific discipline. In: Medical informatics, e-Health. Springer, Berlin, 2014, pp 1–10

Zeng Q-L, Li D-D, Yang Y-B (2013) Vikor method with enhanced accuracy for multiple criteria decision making in healthcare management. J Med Syst 37(2):1–9

Chapter 18
OR Applications in Pharmaceutical Supply Chain Management

Abbas Ahmadi, Mohammad Mousazadeh, S. Ali Torabi, and Mir Saman Pishvaee

18.1 Introduction to Pharmaceutical Supply Chain Management

In order to reach the customers, pharmaceutical products move through multiple players which constitute a supply chain (SC). The complex and dynamic nature of relationships and competition among these members as well as the uncertainty at different levels of pharmaceutical supply chains (PSCs) necessitates the use of advanced while efficient optimization techniques to provide a great suite for informed decision making. Therefore, it is essential to have a solid understanding about the pharmaceutical industry and the mechanisms by which this industry works. This section aims to provide useful information about the context of pharmaceutical supply chains as well as key characteristics of this industry.

18.1.1 Characteristics and Importance of Pharmaceutical Industry

Sustainable development of a nation greatly depends on the health of people live in. Although the pharmaceutical industry has been identified in the United

A. Ahmadi • M.S. Pishvaee
School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran
e-mail: ab_ahmadi@ind.iust.ac.ir; pishvaee@iust.ac.ir

M. Mousazadeh • S.A. Torabi (✉)
School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran
e-mail: mousazadeh@ut.ac.ir; satorabi@ut.ac.ir

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_18

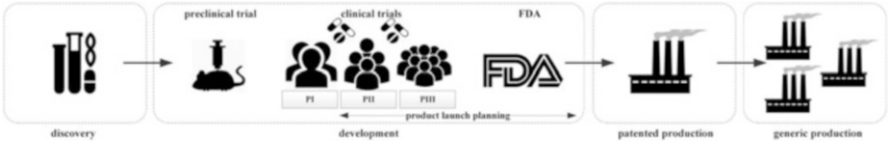

Fig. 18.1 The life cycle stages of a pharmaceutical product (Inspired by Sundaramoorthy et al. (2012a))

Nations' Millennium Development Goals as a major driver for the healthcare sector (Narayana et al. 2012), pharmaceutical companies are still significantly far from effectively managing their resources and satisfying market demands (Mousazadeh et al. 2015; Papageorgiou et al. 2001; Shah 2004). On the other hand, growing non-epidemic diseases (e.g. blood pressure, diabetics and obesity), which are based on unhealthy lifestyle, alongside the discovery of strains of viral infections have put pressure on the pharmaceutical industry (Narayana et al. 2012). To reduce this pressure, pharmaceutical companies look forward to exploit economies of scale in manufacturing and improving the management of their resources such as facilities, materials, human, information, and finances. As an immense global industry, Shah (2004) defines the pharmaceutical industry as "a complex of processes, operations and organizations involved in the discovery, development and manufacture of drugs and medications".

According to Sundaramoorthy et al. (2012a), pharmaceutical products' lifecycle consists of four stages including: discovery, development, patented production, and generic production (see Fig. 18.1). Upon the successful examination and investigation of a new product/compound (i.e. discovery stage, which has a low success rate of less than 1 per 1000 (Tollman et al. 2011)), it is proceeded to the development stage. At this stage, firstly the compound is tested on animals in order to check its safety (i.e. preclinical trial), then a series of clinical trials are applied at different scales (usually in three phases). Here, the success rate is approximately 1 per 5 (Tollman et al. 2011). This stage is fulfilled by USA Food and Drug Administration (FDA) approval. If a failure occurs at any point of the development stage, all the efforts will be futile. It is noteworthy that the first two stages usually take long times and consume huge amount of money (e.g. 8–15 years and even more, and over $2 billion). If a given product passes the development stage successfully, it is introduced commercially into the associate market under patent protection (i.e. patented production stage). This period is called *effective lifecycle*. Afterwards, the product becomes a generic one and is produced by other companies (i.e. generic production stage).

The pharmaceutical industry is generally based on batch production, in which the products are produced through different production stages/manufacturing pipelines, by employing multi-purpose equipment. In the case of shifting the production line from a product to another one, for the sake of purifying and strict legal regulations, a heavy and time-consuming cleaning process (e.g. 2 weeks) is required (Amaro and Barbosa-Póvoa 2008; Caillet 2015; Grunow et al. 2003; Naresh 2012). Since

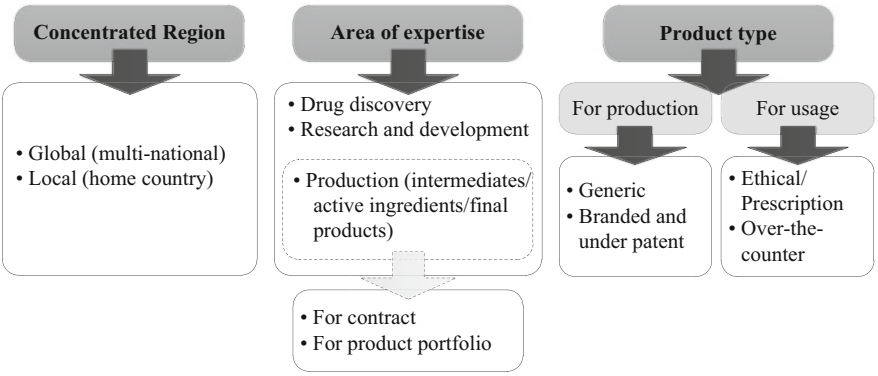

Fig. 18.2 The classification of PSC's characteristics

the cleaning operations and detailed quality control of each batch are very time consuming and also cost-intensive, the companies adopted an enhanced production strategy known as campaign mode. In this method, after doing a family setup, a batch of each product belonging to this family is produced while the available setup remains unchanged for a number of successive batches in this family. Hence, each product is executed only in one or a few campaigns each year (Grunow et al. 2003; Papageorgiou and Pantelides 1996; Naresh 2012).

In order to provide a good insight about the characteristics by which pharmaceutical companies could be distinguished, a classification of various aspects is presented in Fig. 18.2. As a result of global competition raised in all industries (including the pharmaceutical industry) nowadays, pharmaceutical companies may have manufacturing sites in one or many locations around the world. The core competency of these companies may include the discovery of new drugs, the research on different products that are produced elsewhere (generally on the branded products), and the production of various products in the industry (including intermediates, active ingredients, and final products). A company may produce for its own product portfolio or for a third party company through a contract. The products can be categorized into the two main groups of (1) "for production", and (2) "for usage" in terms of type (Shah 2004).

18.1.2 Pharmaceutical Supply Chain

18.1.2.1 Importance and Drivers

The pharmaceutical industry is a large part associated with healthcare sector; such that it carries a considerable portion of the healthcare expenditures. For example, Kelle et al. (2012) stated that it included about 10% of annual healthcare expenses

in the US and nearly $600 billion globally in 2009. However, nowadays, different reasons such as (1) escalating in the costs of introducing new products, (2) declining in the productivity of the research and development (R&D) activities and (3) shortening the effective lifecycle of the products, lead to serious difficulties in recovering the investments has been made in this field. Moreover, due to the emergence of the diseases caused by unhealthy lifestyle as well as aging population, it is a global experience over increasing growth in demands for pharmaceuticals which then puts high pressure on prices and prescription policies and subsequently on pharmaceutical industry (Shah 2004; Sousa et al. 2011; Gatica et al. 2003). On the other hand, over the time, factors such as expiry of short-life patents, the emergence of generic substitutes, etc. have resulted in the shortening of the effective lifecycles of new products (Gatica et al. 2003). Also, the competition among companies has altered the competition among the supply chains even in the global level.

Although great advances and improvements have been achieved in the manufacturing, storage, and distribution systems, but several pharmaceutical companies suffer from the inability of efficiently satisfying market demands. Therefore, the pharmaceutical supply chains are potentially ready to benefit from the efficient optimization techniques. In summary, supply chain management is an accepted methodology to protect the pharmaceutical industry against the introduced pressures, to increase the profit margins and to overcome the obstacles of efficiency addressed previously, while taking into consideration their limited available resources (Mousazadeh et al. 2015; Shah 2004; Papageorgiou 2009; Masoumi et al. 2012).

18.1.2.2 Characteristics and Scope

Following the wide definition of WHO (world health organization) for the drug or pharmaceutical preparation (Shah 2004), a PSC consists of many components, generally including suppliers, primary and secondary manufacturing sites, waste plants, distribution centers, wholesalers, and retailers, as it is well-illustrated in Fig. 18.3. Pharmaceutical products are generally produced in two stages, i.e. the primary and secondary manufacturing stages. Being more detailed, the primary manufacturing site transforms raw material into active pharmaceutical ingredient (API) through various chemical syntheses and separation stages or throughout fermentation, product recovery and purification. Then, the API is converted to final product/drug (e.g., tablets, solutions, vials, inhalers, etc.) by adding varieties of excipients at the secondary manufacturing site. Typically, secondary plants outnumber the primary ones. Thereafter, the bulk quantities of final products are packaged in different sizes in main DCs, in which a great storage capacity is available and then the packages are delivered to wholesalers/local distribution

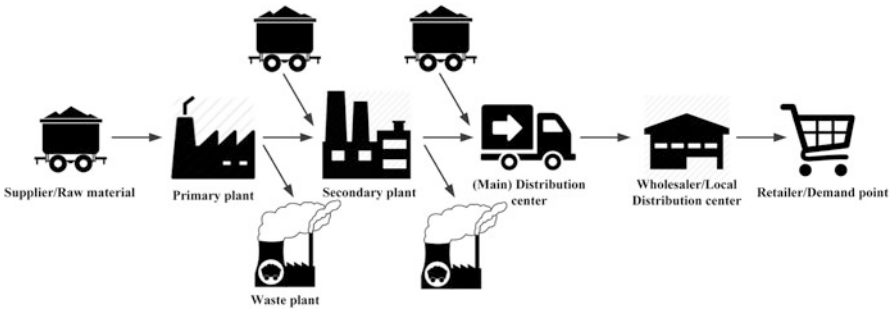

Fig. 18.3 A schematic view of a pharmaceutical supply chain (PSC)

centers (DCs). Finally, the local DCs are responsible to dispense the drugs among demand points (e.g., pharmacies, hospitals, clinics, etc.) which are in contact with the patients. It should be noted that in addition to primary plants, the secondary plants and DCs may also utilize external suppliers. Moreover, the wastes released over the supply chain are managed by waste plants (Mousazadeh et al. 2015; Sousa et al. 2011; Levis and Papageorgiou 2004; Burns 2002b; Laínez et al. 2012; Susarla and Karimi 2012).

In an integrated manner, Burns (2002a) categorized all PSC's players into three main groups, i.e. producers, purchasers and pharmaceutical providers. "Producers" refers to all pharmaceutical companies, medical surgical product companies, device manufacturers, and manufacturers of capital equipment and information systems. Also, "purchasers" refers to grouped purchasing organizations (GPOs), pharmaceutical wholesalers, medical surgical distributors, independent contracted distributors, and product representatives. Finally, "pharmaceutical providers" are all hospitals, hospital systems, integrated delivery networks (IDNs), and alternative site facilities.

The primary and the secondary plants are the key players of a PSC, which constitute the SC backbone and affect the framework generally. The remaining echelons do not significantly influence the structure of a PSC. The aforementioned two echelon production system (i.e. two stages of producing APIs and final products), which is the common adopted approach in the literature, is suffering from some inefficiencies such as (1) high cost of inventory, logistics, and production; (2) long supply chain cycle times; and (3) uncompromising supply chain (i.e., an inflexible supply chain which is not able to be resilient against risks). Therefore, the Novartis-MIT Center for Continuous Manufacturing has started a novel project to change the production technology from batch mode to the continuous mode, where raw material from the APIs to the final products are processed continuously in a single integrated facility. The results achieved from implementing this new integrated end-to-end production strategy (see Fig. 18.4) show a cost saving of around 9–40%, while ensuring a quick response of supply chain to the demand changes (Sundaramoorthy et al. 2012a; Schaber et al. 2011).

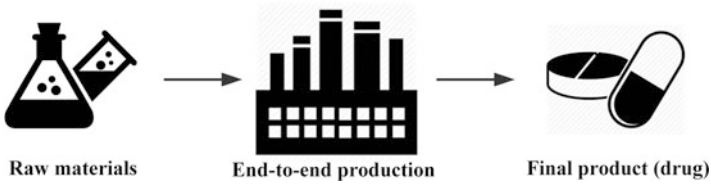

Fig. 18.4 A Schematic view of continuous production mode (Adapted from Sundaramoorthy et al. (2012a))

18.1.2.3 Differences of PSCs with Other SCs

Laínez et al. (2012) highlighted several features of PSCs which differentiate them from other typical SCs. These differences could be summarized as follows:

- Time-consuming, high engagement of resources and low success rate at product discovery stage (see stage one in Fig. 18.1). This fact is due to the limited knowledge about various diseases and medical conditions as well as the extreme uncertainty involved in the process.
- The long period of extremely costly clinical trial (see stage two in Fig. 18.1)
- Extreme regulatory burdens, because of the direct effect on the health and safety of society.
- The short shelf life and perishability nature of the products (drugs).
- Extremely long production cycle times (e.g. up to 9 months).
- The presence of high uncertainty in demand.
- The extended geographically network of PSCs and specific features of raw material and products; which altogether may impose the employment of cold supply chains.
- The short period of effective lifecycle and the generic competition at the end of this period.

The issues of PSC management cover several decisions in strategic, tactical, and operational levels. Therefore, the pharmaceutical companies are quite ready for receiving help from effective optimization techniques in order to improve their performance (Mousazadeh et al. 2015). Adopting these optimization techniques will also help companies to snatch more market share in today's competitive local and global markets. This chapter generally focuses on identifying the decision-making problems in the PSC and the associated mathematical programming approaches to model the problems as well as conducting the decisions in an optimal direction. The next section is dedicated to introducing the decisions which must be made at different levels of a PSC and their modeling approaches. Then, the related solution methodologies are identified in Sect. 18.3. Section 18.4 provides a real-life example for a typical PSC design and several important future research avenues are presented in Sect. 18.5.

18.2 Classification of Problems and Challenges

This section seeks to identify different problems prevalent in PSCs as well as classifying them in appropriate categories according to the different decision making time horizons. In order to optimize the decisions, two or three levels of decisions should simultaneously be included in the model, instead of optimizing them independently. Hence, in the following, in order to provide sufficient insight, a subsection is dedicated to each decision level followed by a discussion on integrated approaches. Figure 18.5 provides an illustration and a summary of the literature.

18.2.1 General Aspects

In the literature, many assumptions have been adopted to model the PSCs related decision problems, which are elaborated in the following subsections. Nevertheless, as a general assumption, the majority of the researchers has only considered a single-period decision horizon (Susarla and Karimi 2012; Hansen and Grunow 2015) in their modeling, while models covering a multi-period horizon are so limited (Mousazadeh et al. 2015; Sousa et al. 2011; Susarla and Karimi 2012). The multi period models are mostly based on a discrete representation of time with periods of predefined length, while Grunow et al. (2003) employed continuous representation of time. Also, another assumption which may limit some frameworks in the real-life settings is to include only one product in the model. Recently, researchers have attempted to release this assumption (see (Mousazadeh et al. 2015)) in order to provide more applicable (i.e. multi-product) models.

With respect to the planning horizon length, various decisions may raise at a PSC including (Papageorgiou 2009):

- Strategic/long-term decisions; i.e. (1) number, size and location of primary and secondary production centers, main distribution centers, local distribution centers and other intermediary warehouses, (2) network connectivity including allocation of suppliers to manufacturing sites, manufacturing sites to DCs/warehouses and finally DCs/warehouses to markets, (3) transportation decisions such as modes of transportation between successive stages of the supply chain (e.g., road, rail, sea, etc.) and sizes of material shipments;
- Tactical/mid-term decisions; i.e. (1) replenishment policies, (2) inventory management at main DCs, local DCs and intermediary warehouses, (3) production planning and high-level scheduling of family products at primary and secondary manufacturing centers;
- Operational/short-term decisions; i.e. scheduling/sequencing decisions, including assignment of different product stages to the production lines, numbers and size of batches, length of campaigns, and material flows.

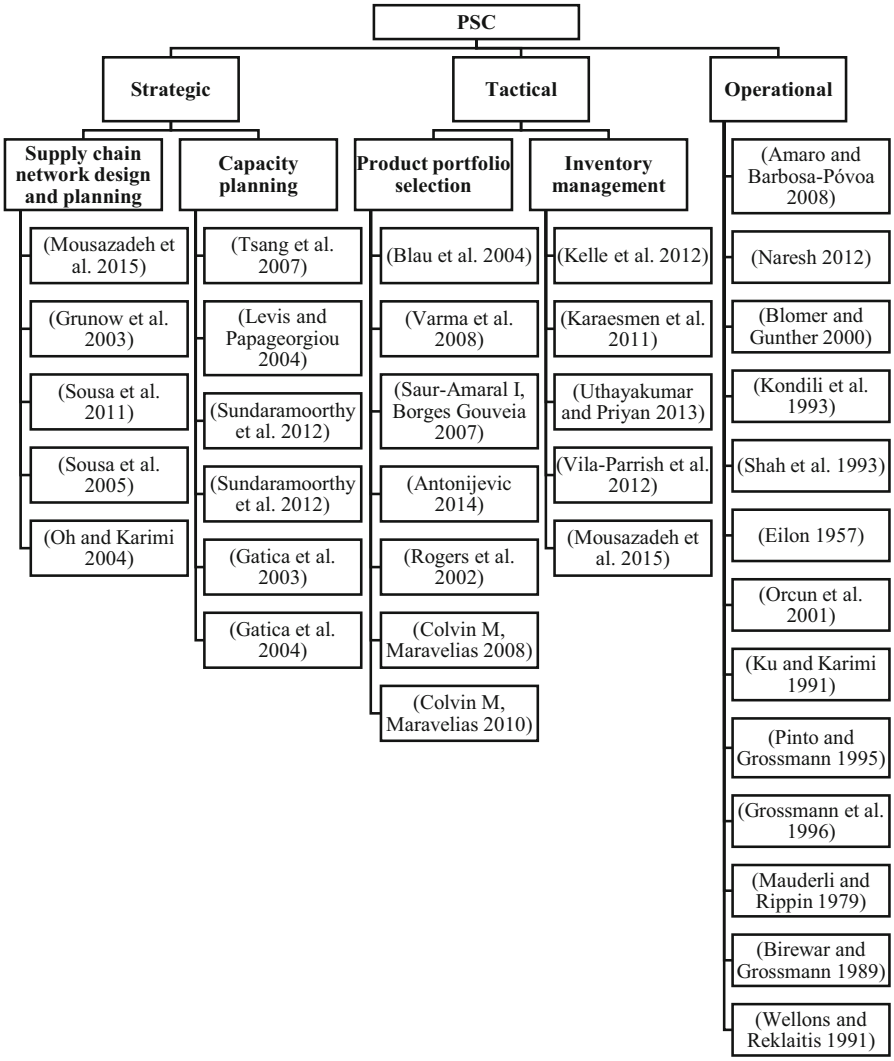

Fig. 18.5 A summary of the literature

Different objective functions may be deployed as a quantitative measure to assess the supply chain. These objective functions can be divided into the two main categories, i.e., (1) financial flow based measures and (2) customer responsiveness based measures. Minimization of costs and inventory investments and maximization of sales, profit and return on investment, as well as minimization of lead time, customer response time, and product lateness, also maximization of fill rate are some examples.

18.2.2 Strategic Decisions

Strategic decisions deal with a long time horizon (e.g. 2–5 years). They commonly require large financial investment and their effects remain over the time and influence the subsequent decisions through imposing several constraints on tactical and operational decisions.

18.2.2.1 Supply Chain Network Design and Planning

Supply chain network design (SCND) refers to determining the optimal network structure (i.e. facility locations and allocations) as well as to determine how to use the production, distribution and storage resources optimally in the whole chain in order to quickly and appropriately respond to customer orders and demands (Papageorgiou 2009). Also, since the primary and the secondary manufacturing plants are often established in scattered locations globally, they often would be assigned to different locations, even different countries as well. Therefore, it is very important to provide a coordinated plan for different echelons of the supply network to achieve an excellent performance for the whole network through integration of supply, production and distribution echelons (Mousazadeh et al. 2015; Shah 2004). This research area has a rich literature in general, but few works have been conducted for the PSCs, despite of its aforementioned special differences.

In producing APIs, due to the nature of low-volume and high-value (LVHV) of products, transportation costs of products are of minor relevance. Hence, the primary plants may be established anywhere around the globe (Sousa et al. 2011). This fact alongside other reasons of global business, often leads to a geographically dispersed supply chain, i.e. global SCs. However, transit times between different plants are necessary to be considered (Grunow et al. 2003). Many criteria have to be taken into account while deciding about the best locations for establishing new facilities, i.e. tax rates (free trade zones may reduce tax rates), the costs of resources, political and economic issues, safety and security concerns, etc. On the other hand, at the end of the PSC, the large bulk of products bolds the transportation costs (Sousa et al. 2011).

In order to mathematically model a PSC network, the researchers adopted different assumptions. Mousazadeh et al. (2015) considered an Iranian PSC by the assumption that the APIs are imported from overseas and disregarded the echelon of primary manufacturers. A multi-period global supply chain network was investigated by Sousa et al. (2005), in which the primary and secondary echelons along with their warehouses, as well as the market place of final products have been included while ignoring the distribution network. They supposed that there are five areas for secondary sites and markets and there is no flow among different areas due to the high transportation costs. It is also assumed that each product is produced in one and only one site (for both primary and secondary plants) at each time period. While, most of the relevant models supposed that at most two different

types of products are produced in each period; transportation of products between different plants is not allowed; and coordination of campaign/batch schedules at different production stages is limited to within a plant, instead of considering various plants (Grunow et al. 2003). The majority of current papers has considered only one objective to be optimized, which reflects the minimization of costs or maximization of net profit value (NPV) (see Eq. 18.1) (Sousa et al. 2005). Whereas, recently Mousazadeh et al. (2015) took into account the minimization of total costs and unsatisfied demands simultaneously, which led to a bi-objective model. Indeed, when investigating a global supply chain (which is the case in the most of PSCs), some issues such as different capital and operating costs, tax rates, duties (Oh and Karimi 2004; Oh and Karimi 2006), exchange rates, transfer prices (Yi and Reklaitis 2007) and other aspects of the problem that influence optimization of the model should be regarded. Since different countries impose various tax rates and with respect to the fact that the profit after taxes may be significantly changed, researchers commonly have considered maximization of NPV as the objective function, instead of directly minimization of costs. Production costs, inventory handling costs, transportation costs, unsatisfied demand costs, tax costs, products allocation costs, and revenues are some examples of the components that can be involved in the objective function.

$$NPV = \sum_{time} (revenues - products\ costs - sites\ costs - tax) \tag{18.1}$$

Different decisions may be incorporated in the model, such as numbers, locations and production technologies of manufacturers plants; numbers, locations and capacities of main distribution centers; numbers and locations of local distribution centers (Mousazadeh et al. 2015); as well as location and allocation of secondary and primary sites and also products (Sousa et al. 2005).

In order to provide a realistic model, the popular constraints that should be taken into consideration (according to the problem under consideration) are as follows: primary and secondary product allocation, which guarantees that each product in the portfolio is allocated at least (or exactly) to one manufacturing site; capacity constraints, which limit the use of resources (including equipment, materials, etc.) at their available level; constraints related to balance of flows between different echelons of the supply chain (e.g. between markets and warehouses, warehouses and manufacturing sites, etc.); inventory constraints establishing the relationship between inventories and flows; constraints for size of campaigns imposing the number of batches allowed to be included within a campaign (in campaign production mode); the amount of products that can be produced in each batch (i.e. batch size); time duration allocated to each campaign and/or batch; the capacity of resources; appropriate sequence of tasks; material flows between sites, batches, campaigns, etc.; demand and required quantities of final products; so on.

In this respect, Grunow et al. (2003) considered the supply network planning problem of a global pharmaceutical supply chain in which each product might be produced in more than one country. Their formulation is benefited from the

continuous representation of time to simplify the developed mixed integer linear programming (MILP) model in terms of the formulation of the maximum and minimum campaign length constraints, which are the main concerns in the highly regulated pharmaceutical industry. Therefore, a model with high efficiency in terms of computational time was obtained.

A typical structure of the pharmaceutical supply chain network alongside its mathematical model presented by Mousazadeh et al. (2015) is proposed in "case study" section.

18.2.2.2 Capacity Planning

The time to market is the most important issue in the pharmaceutical industry. This is due to the fact that the sustainable growth of pharmaceutical companies strongly depends on "blockbuster" drugs. Therefore, the new products must be introduced into the market as early as possible, which necessitates the strategic planning for making available the required and enough capacities to fulfill demand (Sundaramoorthy et al. 2012a; Laínez et al. 2012).

Since the PSC is generally a global SC, multi-site capacity planning is often required. Indeed, the industry aims to integrate traditionally isolated areas such as product development, manufacturing, accounting and marketing (Levis and Papageorgiou 2004). The decisions associated with such issues are taken into consideration during the development stage (see Fig. 18.1).

The most prevalent restrictive assumptions, which have been applied in the extant models in the literature are as follows: unlimited raw material supply, known demand forecast and prices for new and existing products, and limited time horizon. While, the decisions are affected by data uncertainty that commonly arises from the indeterminate nature of clinical trials and also demand side. Taking these facts into account, Tsang et al. (2007) proposed a stochastic multi-stage, multi-period and multi-scenario mixed integer optimization model for multiple vaccines in which demand was considered as uncertain parameter and the well-known scenario-based approach was used to handle this type of uncertainty. Also, various expected clinical trial outcomes with their probabilities of occurrence were identified, while the details of processing times, investment costs and production scale were also taken into account in the proposed model. The expected NPV was used as the objective function to assess the decisions as well.

In another work, Sundaramoorthy et al. (2012a) developed a framework to tackle with the challenges prevalent in capacity planning for the novel integrated end-to-end production scheme based on continuous manufacturing mode (see Fig. 18.4) in the presence of uncertainties. It is worthwhile to go into the details of this framework. Suppose we are dealing with a pharmaceutical company having a portfolio of products undergoing development process. Each product has passed its PI trial and should be investigated in the later trials (i.e. PII and PIII) in series. It is assumed that the probability of success (or failure) for each trial, demand trends during product lifecycle, price and costs are known in advance. If the product

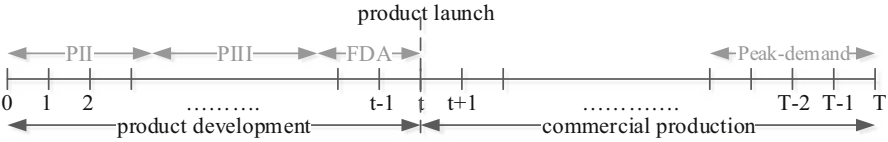

Fig. 18.6 Planning horizon from product development to commercial production (Adapted from Sundaramoorthy et al. (2012a))

(drug) successfully passes both the trials and receive the FDA approval, it will be ready for commercial production. The company decided to perform in the integrated production scheme, by having a set of production facilities that should be set up either in dedicated or multipurpose state. The planning horizon is divided into T time periods, which starts from the beginning of the PII trial, followed by PIII and FDA periods and finally ends to the peak-demand periods at the product lifecycle, which includes two main activities of product development and commercial production (see Fig. 18.6).

In this problem, decisions include (1) the best time for establishing a new facility or expanding capacity of an existing facility, (2) the optimum capacity of the newly established facilities and/or optimum expansion in the capacity of existing facilities, (3) the portfolio of products that must be produced in each newly established/existing facility, (4) the production volume of each product in each time period and finally, (5) inventory level of each product at the end of each time period. This problem was formulated as an MILP model based on the two-stage stochastic programming approach. In this framework, the first two sets of decisions are considered as here-and-now decisions and the others as wait-and-see decisions (these terms are discussed in more details in Sect. 18.3).

18.2.3 Tactical Decisions

Tactical decisions concern with the mid-term planning of all activities, which happen through the product lifecycle ranging from operations requiring the clinical trials to delivering the drugs to the end users. Detailed product flows between facilities of different echelons (transportation) and at the same echelons (transshipment (Chiou 2008)) of the PSC, transportation modes, production rates in manufacturing plants, inventory of products, products flow from sites to markets, product sales, key resource allocations, new product introductions, maintenance, material transfers, product portfolio selection, and changeovers in an integrated manner for a plant are some of the decisions to be made at this level (Mousazadeh et al. 2015; Naresh 2012; Oh and Karimi 2004). Considering any combination of the aforementioned decisions would generate a different model.

The existence of multi-stage production systems and multi-purpose equipment in the pharmaceutical industry has faced the mid-term planning of PSCs with complexity; especially when material flows between different production locations.

Therefore, providing a coordinated procurement, production and distribution plan within the entire supply network would be valuable, in order to avoid sub-optimality while unnecessary inventories, inefficient capacity utilization, uneconomical allocation of resources, etc. (Grunow et al. 2003).

Amaro and Barbosa-Póvoa (2008) presented an MILP model to incorporate the decisions related to the production, customization, distribution, transportation and recycling/remanufacturing in a closed loop global PSC, in which the reverse flows are accounted for and the operations, resources and market places are geographically dispersed. They divided time horizon into a number of discrete time intervals wherein the related decisions must be made. They assumed that all the SC activities such as transportation and processing ones have a fixed completion time, while in the real-world it does not take place. Also, a set of linear constraints, describing the core of SC operations (topology, operational network, material and markets) were implemented. The model aimed to maximize the global SC profit, as well as to support economical criterion.

18.2.3.1 Product Portfolio Selection

If a firm wants to maintain its competitive position in the market, it must introduce new products to the market continuously. In this way, one important challenge is that what products should be included in the R&D portfolio (Blau et al. 2004) and how to allocate resources (Varma et al. 2008) to achieve the greatest possible return on investment in a fast and reliable way. Such decisions have to be made under considerable uncertainties in the PSCs (such as uncertainties in demand, sales price, outcomes of clinical trials, risk of the failure or prolongation of getting the success in R&D process, emergence of an alternative product, etc.) (Gatica et al. 2003; Traple et al. 2014; Saur-Amaral and Borges Gouveia 2007; Antonijevic 2014). R&D costs associated with the development of the new product, and the product's commercial characteristics such as manufacturing costs, selling prices, marketing expenses are several significant features of a product that can be considered for deciding about this matter (Papageorgiou et al. 2001; Rogers et al. 2002). Also, the constraints associated with capital budgeting, resources, technical and regulatory limitations are commonly imposed to the problem.

Among the different methods for handling uncertainty, stochastic programming (Colvin and Maravelias 2008, 2010) is the most common method deployed for approaching the uncertainty in portfolio selection. As an example, Rogers et al. (2002), proposed a continuous decision making method to address the problem related to development of a new product. This method was composed of a series of continuation/abandonment alternatives at each stage of R&D process to decide whether to proceed further or stop the development efforts. They considered four outcomes, i.e. (1) high success, (2) target success, (3) low success and, (4) failure, each of which with predefined estimated probabilities for clinical trials. Then, a multi-stage stochastic programming approach in the form of an MILP formulation was extended. A more detailed description of uncertainty handling methods is provided later in Sect. 18.3.

18.2.3.2 Inventory Management

Owing to the direct impact of PSCs on health and safety, a high customer service level is necessary. Hence, many pharmaceutical industries try to meet a fill rate close to 100% and prevent to encountering shortage, by providing a large amount of inventories in their SCs. In addition to the high capital costs beyond this strategy and also the challenges related to storage space limitations, PSCs suffer from the perishability of products, which annually imposes millions of dollars as financial losses (Karaesmen et al. 2011; Uthayakumar and Priyan 2013).

The aforementioned challenges reveal the need for the appropriate management on the inventory issues such as order quantities, purchasing dates, inventory level alongside considering customer satisfaction without negatively affecting the public health, patient safety, or relationships with the PSC members (Uthayakumar and Priyan 2013; Vila-Parrish et al. 2012). In spite of these facts, this area was out of attention in the past, and has recently received more attention from researchers in the field of operations research (OR).

Among the inventory policies extended in the inventory management literature, some policies such as the periodic-review is not acceptable for industries related to public health (due to the risk of shortage). Therefore, the common approach adopted for PSC is the continuous-review. Due to the particular nature of the pharmaceutical products (e.g. high prices and costs of shortage, the presence of sensitive molecular entities degrading over time, etc. (Laínez et al. 2012)), compared to other perishable products, they do not comply the models prevalent in the literature. In summary, these issues place serious limitations on inventory management and other imposed risks to the enterprise. Hence, the inventory management in PSCs requires particular models to be established while taking into account all of the specific aspects.

In this respect, Uthayakumar and Priyan (2013) proposed a model, based on the continuous-review policy, for a two-echelon PSC with multiple products. The PSC involved a pharmaceutical company and a hospital. Their models covered delay in payments, customers' satisfaction, lead times (which could be controlled via crash cost), a maximum allowable inventory costs, and space availability. The model aimed to determine the optimal inventory lot sizes, lead times, and the total number of deliveries from the pharmaceutical company to the hospital in a production cycle to minimize the total cost of the PSC and meet the target hospital customer service level and its space availability.

Kelle et al. (2012) focused on inventory management of a PSC of a local hospital. They took into consideration three different parties affecting the performance of the supply chain, including the: physicians, pharmacists, and a group purchasing organization (GPO). A basic conflict between the parties was highlighted, in the viewpoint of the product variety versus economies of scale. Indeed, physicians tend to have a large variety of drugs on hand to perfectly meet the goals in the treatment of the patients. In contrast, the pharmacy directors and GPO strive to reduce the diversity in order to achieve better prices due to economies of scale. This situation provides a complex situation to model the problem, which necessitates the use of multiple models, instead of a single one. This issue is due to the fact that the related

decisions cannot be ordinarily made using a hierarchical top-down or bottom-up framework. Hence, to tackle with the issue, they presented a framework in which the top level decisions were firstly determined as the current condition; after which the decisions of the lower level were optimized. Afterwards, an iterative top-down and bottom-up procedure was provided to find the solution having the best effect on the overall performance of the system. Also, in order to handle the change and uncertainty in demand, a very costly emergency delivery takes place in the case of shortage.

18.2.4 Operational Decisions

Operational decisions specifically address short time-scales, e.g. daily or weekly, and involve more details compared to the strategic and tactical decisions. The models provided for the operational level may be used for other levels; but due to the details incorporated, solving them for long time horizon would be an intractable task (Naresh 2012).

Due to the low manufacturing speed at primary and secondary plants, the PSC cycle time may strangely be lengthy (e.g. 300 days) (Shah 2004). This fact reveals the need for high quality decisions at the operational level, to provide the capability of quick response to demand as well as the changes. Scheduling problems are the most concentrated area in this decision level. Although, it may be included within the tactical decisions under longer time horizons (Naresh 2012).

Scheduling /sequencing encompasses a variety of short-term decisions ranging from assignment of different product stages to the production lines, numbers and size of batches, length of campaigns, cleaning and set-ups, material flows, storage, etc. Different combinations of these decisions generate a large number of problems (Naresh 2012).

Given the batch production nature of the pharmaceutical industry (especially in primary plants, which are mainly based on chemical processes) and the presence of multiple stage processes and multi-purpose equipment, we face with complex scheduling problems for production purposes at the operational level (Naresh 2012).

The scheduling concept has a rich literature and has been studied for many years ago from different perspectives (Blomer and Gunther 2000; Shah et al. 1993; Eilon 1957; Orcun et al. 2001) and using different approaches (such as metaheuristic algorithms (Ku and Karimi 1991; Jou 2005; Shelokar et al. 2004) and mathematical modelling methods (Pritsker et al. 1969; Bowman 1959)). However, this matter has received little attention in the pharmaceutical industry. Scheduling in pharmaceutical plants can be divided into two main categories (according to the operating conditions): batch scheduling and campaign scheduling (Naresh 2012).

Chemical batch processes are generally represented by state-task-network (STN) or resource-task-network (RTN) techniques having a network-like structure. These representations are helpful to model the scheduling problems. Different MILP models (Amaro and Barbosa-Póvoa 2008; Pinto and Grossmann 1995; Blömer

and Günther 1998; Grossmann et al. 1996) have been formulated to deal with this matter (see (Méndez et al. 2006) for a review). Batch scheduling problems can be investigated according to time horizon that can either be discrete (Kondili et al. 1993) or continuous (Orcun et al. 2001). In the discrete approaches, the time horizon is divided into a number of fixed length intervals. If the number of these intervals is large (particularly in production of APIs possessing long time intervals, due to the large number of production stages), the precision is increased, but on the other hand it leads to increase in the complexity of the problem. While, in the continuous approaches, the length of time intervals is not fixed and are considered as variable.

Since, in the batch processing mode, the set-up time for shifting the production line from a product to another one spends a lot of time (due to the cleaning procedures), an alternative strategy is to produce more than the required amount (i.e. long product campaigns). However, it results in high inventory levels. So, campaign scheduling tries to provide a trade-off between inventory cost and set-up times. In this respect, various formulations such as MILP (Mauderli and Rippin 1979; Shah and Pantelides 1991), NLP (Birewar and Grossmann 1989) and MINLP (Wellons and Reklaitis 1991) have been presented so far.

18.2.5 Integrated Approaches

New trends tend to aggregate different levels of decisions and take a cross-functional and coordinated approach, instead of focusing on sub-decisions which lead to fairly poor results. For example, Mousazadeh et al. (2015) took into consideration strategic and tactical decisions concurrently in a network design problem. Indeed, considering the more decisions related to a PSC simultaneously, the more strong results would be obtained. However, this scheme makes the problem more complicated in terms of modeling and particularly obtaining its solution.

Maravelias and Sung (2009) reviewed the integration of scheduling and planning decisions and the methodologies presented in the literature for this purpose. Also, they introduced the existing challenges and opportunities of the work. In this respect, Amaro and Barbosa-Póvoa (2008) aggregated tactical and operational decisions related to planning and scheduling in SC, respectively. They adopted a sequential approach, in which after dividing the time horizon (e.g. 3 months) into some discrete time intervals (e.g. 1 week) and obtaining planning results, a scheduling problem was run for each of the time intervals in smaller scales (e.g. 2 h).

Susarla and Karimi (2012) took into account the integration of production planning, procurement, distribution, and inventory management beside the effects of international tax differentials, material shelf-lives, waste treatment/disposal in a multinational pharmaceutical enterprise through multi-period horizon. Instead of presenting a complex model and focusing on the development of a solution approach, they proposed a simple yet a holistic MILP model which could be solved quickly and could be implanted easily. A positive feature of this model was that it incorporated different echelons of the supply chain from procurement

of raw materials to distribution of final products. Also, Levis and Papageorgiou (2004) presented a large-scale MILP for maximizing of the expected NPV following concurrently determining the product portfolio and multi-site capacity planning in the presence of uncertainties occurred in the clinical trials, in addition to considering the trading structure of the company.

Rotstein et al. (1999) and Papageorgiou et al. (2001) considered development and introduction strategy of the product, as well as capacity planning and investment strategy. Rotstein et al. (1999) formulated the problem as an MILP two stage stochastic programming model. At the first stage, the decisions related to product selection, initial capacity investment, and the initial allocation of manufacturing resources to products are established as "here-and-now decisions". The second stage reflected the decisions of type "wait-and-see" such as additional capacity investment in the case of favorable outcomes, abandonment/sale of capacity for unfavorable outcomes, re-allocation of manufacturing resources to products, and production plans. Indeed, at the second stage different combination of scenarios obtained from the first stage were investigated. The objective function was calculated as the expected NPV using scenarios probabilities. Papageorgiou et al. (2001) also formulated the aforementioned problem as an MILP model through maximizing the NPV as the objective function.

Gatica et al. (2003) examined the multi-stage and multi-scenario case where, using the probability approach, different outcomes that may occur during the development stage and introducing phase of a new drug were incorporated. Their main decisions in the proposed model could be listed as: final portfolio of products, capacity planning, the most appropriate production plans for each production line, product allocation and amounts of products to be manufactured as well as sales and inventory planning. They developed a multi-scenario MILP model consisting the sets of constraints such as production line availability, manufacturing, material balance, lifetime, sales, inventory level, and product selection.

Sundaramoorthy and Karimi (2004) considered the pharmaceutical plants operating in campaign mode and having multiple parallel production lines. They presented an MILP model for the problem of introducing new products to the existing manufacturing facilities beside determining the production planning of (new and old) products, as well as inventory management and sequencing of campaigns, under continuous-time and multi-period horizon.

18.3 Solution Approaches and Challenges

This section deals with the ways and methods of finding the solution of various problems introduced in the last section, besides the challenges appeared because of the unavailability of required data (i.e. data uncertainty).

18.3.1 Finding Solution

Different solution methodologies have been deployed to solve developed models in the context of PSCs. For example, in a multi-objective network design problem, Mousazadeh et al. (2015) used ε-constraint method to achieve the final compromise solution and also to provide trade-off analysis between the conflicting objectives in the MILP problem. In addition, the well-known TH approach (Torabi and Hassini 2008) was employed to obtain a compromise solution in accordance with the decision maker preferences. Gatica et al. (2003) utilized the branch and bound technique through CPLEX solver to find the solution of the formulated MILP model.

As we know, the majority of the mathematical models developed for the real-world problems are intractable in solution procedure. Hereupon, the researchers attempted to employ different strategies (such as decomposition, heuristic methods, etc.) to handle the problems in large-sizes.

Decomposition method is a well-known technique to reduce the burden of computational efforts. It follows two principle concept of (1) decomposing the problem into smaller sub-problems (with the purpose of reducing size of the main problem), and (2) elimination of complicating constraints (Sousa et al. 2011). In the related literature, the approaches such as bi-level decomposition algorithm (Iyer and Grossmann 1998; Bok et al. 2000), Benders decomposition approach (Üster et al. 2007), and Lagrangean decomposition method (Gupta and Maranas 1999) have been adopted in order to solve the large and complex MIP models resulting from the supply chain optimization problems. In this respect, Sousa et al. (2005) adopted a Lagrangean decomposition method to solve the extended MILP model, through utilizing the intrinsic structural matrix of the model. In addition, a heuristic algorithm named PFA (product frames algorithm) was introduced, in which firstly the binary variables related to the locations are sequentially determined instead of simultaneously. Sousa et al. (2011) proposed two decomposition methods for the MILP presented model. The idea beyond the first one was to separate the SC into two main echelons including primary and secondary manufactures respectively, while in the second method, the model was decomposed into several independent sub-problems (one per each time period). Sundaramoorthy et al. (2012b) provided a framework using non-convex generalized Benders decomposition (NGBD) method to solve the problems in continuous pharmaceutical manufacturing having millions variables and constraints within a few hours (e.g. 6 h). As previously mentioned, considering more details about the problem by Amaro and Barbosa-Póvoa (2008) lead to a more complex problem. Therefore they adopted a common industrial practice in which the problem was divided into different levels of details, namely planning and scheduling; by this feature that firstly the planning decisions were made in a time horizon and then further details were investigated in the shorter time domain. Finally, the developed MIP model for each level was solved using CPLEX solver. Also, Rotstein et al. (1999) presented an algorithm based on a hierarchical procedure to solve industrial-case examples of an MILP scenario-based model.

In heuristic techniques, various ideas have been expressed. For example, Susarla and Karimi (2012) firstly solved a relaxed version of the MILP model and found that many of the integer variables in a set of variables are zero. Therefore, they supposed that these variables would remain zero in the optimal solution. Then the problem was solved using CPLEX solver in a quick manner with a little compromise on solution quality. Levis and Papageorgiou (2004) proposed a hierarchical algorithm based on the decoupling of different decision-making levels have been taken into consideration in the SC. Indeed, at the first step, the variables corresponding to the higher level problem were determined and then were fixed at the second level in order to solve the detailed model. A Lagrangean multiplier algorithmic approach was presented by Uthayakumar and Priyan (2013) to obtain the solution of the nonlinear model of inventory problem.

As has been mentioned in the previous section, the discrete representation of time horizon will result in the difficulty of computation (due to the large number of time intervals). In order to obviate this challenge, the researchers tried to examine time temporal aggregation schemes (Grunow et al. 2003), which reduce the number of decision variables and constraints through techniques such as fixing some decision variables. These methods include using micro-periods of scheduling within macro-periods of planning (Wilkinson et al. 1995), employing the concept of rolling time windows (Dimitriadis et al. 1997) and reverse rolling time windows (Bassett et al. 1996), etc.

18.3.2 Uncertainty

Uncertainty in PSCs may arise because of the unavailability, as well as the dynamic and imprecise nature of the required data. In the point of view Dubois et al. (2003), uncertainty can be classified into two categories: (1) Uncertainty in data (which is the most common uncertainty faced in SCs), and (2) Flexibility in constraints and goals. Uncertainty in data is typically described in two forms (Mula et al. 2006): (1) Randomness, that originates from the random nature of the data, and (2) epistemic uncertainty, that is due to the unavailability or insufficiency of needed data, which leads to imprecise data extracted from experts' subjective opinions typically formulated as possibilistic data in the form of triangular or trapezoidal fuzzy numbers. Both stochastic programming and robust possibilistic approaches could be used in order to handle randomness. In detail, the former approach is useful when enough historical data about the random parameters are available, while the latter approach can be deployed in the absence of such historical data. Also, possibilistic programming (Mousazadeh et al. 2015) approaches could be used when dealing with epistemic uncertainty in input data, while flexible programming (Mousazadeh et al. 2014) approaches can deal with flexibility in target values/right hand side (RHC) of soft constraints, respectively. Different classifications of uncertainty in

SCM have been suggested in the literature. Among them, Mousazadeh et al. (2014) presented a comprehensive review on different frameworks used to categorize different types of uncertainty. Also, reviews on uncertainty in process systems and the related optimization schemes can be found in (Sahinidis 2004; Geletu and Li 2014).

In PSCs, the uncertainty may arise in product demand and prices, clinical trials, raw material availability, regulatory changes, investment risk, unit manufacturing and transportation costs, etc. Uncertainty factors prevalent in the PSCs significantly influence the decisions, especially the long term ones. For example, the pharmaceutical companies tend to make a balance between active capacity and anticipated demand in the future, while dealing with very significant uncertainty associated with clinical trials and competitors' activities, emergence of new therapies, etc. (Shah 2004; Laínez et al. 2012). Hence, in order to provide robust decisions under the different uncertain situations (i.e. scenarios), the associated factors should be taken into consideration in a systematic manner (Sousa et al. 2011).

Among the numerous uncertainties mentioned above and the various techniques proposed in the literature to handle uncertainty, limited efforts have been dedicated to dealing with this important issue in the PSCs. Researchers have mainly focused on stochastic and scenario based methods. Indeed, some researchers such as Iyer and Grossmann (1998), while utilizing the stochastic programming, defined a discrete number of scenarios and converted their problem to a multi-scenario optimization model, which behaved in an expectation fashion. For example, Gatica et al. (2003) considered four different scenarios (i.e. high, target, low success, and failure), with given probability distributions, for the purpose of product portfolio selection in the product development process at the primary manufacturing stage. Then, using the decision tree, a multi-stage stochastic problem with the aim of maximizing expected NPV was formulated. This approach gave this opportunity to compare the decisions made in the decision tree, according to their expected NPV. It also provides the probability distribution of NPV to decision makers, to be announced about the risk involved in the project. In another work, the same approach was adopted by Gatica et al. (2001).

A common approach to cope with the stochastic uncertainty is two-stage stochastic programming (Rotstein et al. 1999), in which the decisions of the problem are divided into two sets: (1) the "here-and-now" decisions, which must be made in the first stage and before the realization of uncertain parameters, and (2) the "wait-and-see" decisions, which must be made in the so called "second stage", after realization of all uncertain parameters. In detail, the prior set of decisions must be taken such that the most satisfactory outcome across all possible scenarios could be achieved, while the second stage decisions reflect operational adjustments (i.e. recourse actions) needed after realization of uncertain inputs and

after obtaining new information. This approach has frequently been employed in PSCs (Sundaramoorthy et al. 2012a; Rotstein et al. 1999), as was addressed in the previous sections.

The majority of the published works has accounted for only one type of uncertainty (see (Levis and Papageorgiou 2004; Hansen and Grunow 2015)), while Mousazadeh et al. (2015) dealt with epistemic uncertainty in some of the critical parameters through a robust possibilistic programming approach (see (Pishvaee and Khalaf 2016; Pishvaee et al. 2012) for more information) by adopting a trapezoidal possibility distribution for parameters, in which it is possible for imprecise data to change in pre-defined ranges.

18.4 Case Study

In this section, a real case study presented in Mousazadeh et al. (2015) is reviewed. The authors developed a bi-objective mixed integer linear programming (BOMILP) model for a multi-product multi-period pharmaceutical supply chain network design (PSCND) problem in a national scale which covers several strategic and tactical decisions. It is worth noting that, without losing generality and due to the space limitation, the proposed model in this section is a slightly simplified version to that of Mousazadeh et al. (2015) by limiting it to a single-product, single transportation mode and single production technology.

Consider a pharmaceutical supply chain network (PSCN) consisting of four echelons, i.e. secondary drug manufacturing sites (MSs), main distribution centers (MDCs), local distribution centers (LDCs)/wholesalers and demand zones (DZs). There are some candidate locations for establishing required MSs, MDCs and LDCs. In addition, different possible production capacity levels and storage capacity levels for MSs and LDCs are respectively available. Furthermore, keeping inventory in MDCs and LDCs is possible, however, it is limited to the storage capacity of each center. On the other hand, all the drugs manufactured in each period must be transported to MDCs and cannot be stored at MSs. The demand of each demand zone is only satisfied through the direct shipments from the LDCs. Transshipment from an LDC to another LDC is also considered in order to effectively meet the fluctuation of demands over a dynamic multi-period horizon. In order to meet the emergency needs for a given drug in each period, a minimum safety stock must be kept in each LDC in each period.

The proposed model consists of two contradictory objectives, i.e. minimization of total costs (including the total opening costs, transportation/transshipment costs and inventory holding costs) and minimization of maximum unsatisfied demand in all periods. In detail, establishing more MSs, MDCs and LDCs with more

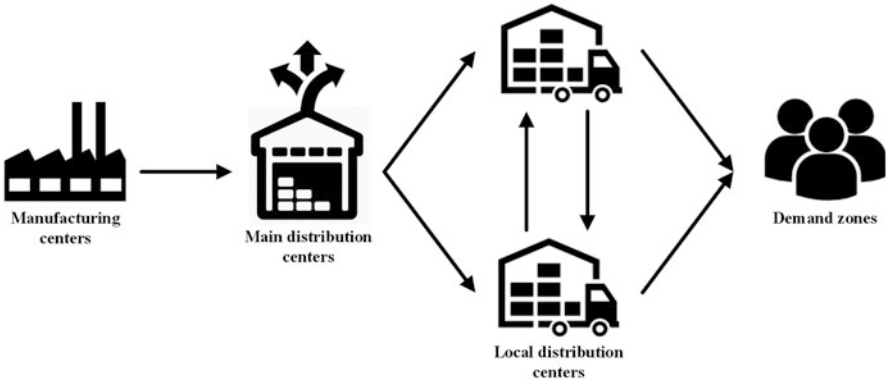

Fig. 18.7 Pharmaceutical supply chain considered in (Mousazadeh et al. 2015)

production and storage capacity, in one hand would increase total costs, but on the other hand would decrease the unmet demand and vice versa. Finally, many strategic decisions, i.e. the location of MSs, MDCs and LDCs as well as some tactical decisions, i.e. optimal material flows over a mid-term planning horizon, holding inventory of each product family in each period, production capacity of MSs and storage capacity of MDCs must be made. The structure of the discussed problem in (Mousazadeh et al. 2015) is replicated in Fig. 18.7.

The indices, parameters and variables used to formulate the concerned PSCN are as follows:

Indices

i	Index of candidate locations for MSs, $i \in \{1, 2, \ldots, I\}$
j	Index of candidate locations for MDCs, $j \in \{1, 2, \ldots, J\}$
k	Index of candidate locations for LDCs, $k \in \{1, 2, \ldots, K\}$
l	Index of demand zones, $l \in \{1, 2, \ldots, L\}$
m	Index of possible production capacity levels for $m \in \{1, 2, \ldots, M\}$
n	Index of possible capacity levels for MDCs, $n \in \{1, 2, \ldots, N\}$
t	Index of periods, $t \in \{1, 2, \ldots, T\}$

Parameters

f_i^m	Fixed cost of opening MS i with production capacity level m
g_j^n	Fixed cost of opening MDC j with capacity level n
h_k	Fixed cost of opening LDC k
c_{ij}	Unit transportation cost from MS i to MDC j
a_{jk}	Unit transportation cost from MDC j to LDC k
e_{kl}	Unit transportation cost from LDC k to demand zone l
$tr_{kk'}$	Unit transshipment cost from LDC k to LDC k'
d_{lt}	Demand of demand zone l at period t
SC_j	Unit storage cost at the end of each period in MDC j
SC'_k	Unit storage cost at the end of each period in LDC k
SS_{kt}	Safety stock in local DC k at the end of period t
δ_j^n	Storage capacity of MDC j established by the capacity level n
γ_k	Storage capacity available at LDC k
τ_i^m	Effective production capacity of MS i with production capacity level m

Variables

I_{jt}	Inventory level at MDC j at the end of period t
I'_{kt}	Inventory level at LDC k at the end of period t
u_{ijt}	Quantity of product manufactured at MS i in period t and is shipped to MDC j
q_{jkt}	Quantity of products shipped from MDC j to LDC k in period t
o_{klt}	Quantity of products shipped from LDC k to DZ l in period t
$v_{kk't}$	Quantity of products transshipped from LDC k to LDC k' in period t
x_i^m	1 if potential MS i with production capacity level m is opened; and 0, otherwise
y_j^n	1 if potential MDC j with capacity level n is opened; and 0, otherwise
z_k	1 if potential LDC k is opened; and 0, otherwise

Using the above mentioned notations, the proposed model is as follows:

$$\begin{aligned}\text{Min } w_1 = &\sum_{i,m} f_i^m x_i^m + \sum_{j,n} g_j^n y_j^n + \sum_k h_k z_k + \sum_{i,j,m,t} c_{ij} u_{ijt} + \sum_{j,k,t} a_{jk} q_{jkt} \\ &+ \sum_{k,l,t} e_{kl} o_{klt} + \sum_{k,l,t} tr_{kk'} v_{kk't} + \sum_{j,t} SC_j I_{jt} + \sum_{k,t} SC'_k I'_{kt}\end{aligned} \tag{18.2}$$

$$\text{Min } w_2 = Max_{l,t}\left\{ d_{lt} - \sum_k o_{klt} \right\} \tag{18.3}$$

s.t.

$$I_{jt} = I_{j,t-1} + \sum_{i} u_{ijt} - \sum_{k} q_{jkt} \quad \forall j, t \tag{18.4}$$

$$I'_{kt} = I'_{k,t-1} + \sum_{j} q_{jkt} - \sum_{l} o_{klt} + \sum_{k'} v_{k'kt} - \sum_{k'} v_{kk't} \quad \forall k, t \tag{18.5}$$

$$\sum_{j} u_{ijt} \leq \sum_{m} x_i^m \tau_i^m \quad \forall i, t \tag{18.6}$$

$$I_{j,t-1} + \sum_{i} u_{ijt} \leq \sum_{n} y_j^n \delta_j^n \quad \forall j, t \tag{18.7}$$

$$I'_{k,t-1} + \sum_{j} q_{jkt} + \sum_{k'} v_{k'kt} \leq \gamma_k z_k \quad \forall k, t \tag{18.8}$$

$$I'_{k,t-1} \geq SS_{kt} z_k \quad \forall k, t \tag{18.9}$$

$$\sum_{m} x_i^m \leq 1, \quad \forall i, \tag{18.10}$$

$$\sum_{n} y_j^n \leq 1, \quad \forall j, \tag{18.11}$$

$$x_i^m, y_j^n, z_k \in \{0, 1\} \quad \forall i, j, m, n, k \tag{18.12}$$

$$u_{iht}, q_{jkt}, o_{klt}, v_{kk't}, I_{jt}, I'_{kt} \geq 0, \quad \forall i, j, k, k', l, m, t \tag{18.13}$$

The objective function (18.2) consists of the fixed opening costs, transportation/transshipment costs and inventory holding costs. The objective function (18.3) tends to minimize the maximum unsatisfied demand in all periods and for all DZs. Constraints (18.4) and (18.5) are the flow balance equations for MDCs and LDCs respectively. Constraint (18.6) ensures that the total production in each opened MS in each period would not violate its respective production capacity. Also, Constraints (18.7) and (18.8) guarantee that the available inventory from the previous period plus the total incoming flows from facilities in previous echelon (shipment) or from facilities in the same echelon (transshipment) cannot exceed the storage capacity of the MDCs and LDCs, respectively. Constraint (18.9) guarantees keeping a minimum safety stock in opened LDCs. Constraints (18.10) and (18.11) guarantee that each MS and MDC could be opened in one of the available production capacity level and storage capacity levels respectively. Finally, constraints (18.12) and (18.13) enforce the binary and non-negativity restrictions on the corresponding decision variables.

In order to verify and analyze the proposed model, it was tested via a real case study in Iran. To do so, they designed a national supply chain network for "amoxicillin 500 mg Cap" product which was the most prescribed and consumed drug among around 5500 different types of drugs in the Islamic Republic of Iran, with around 900 million of taken capsules annually. This high consumption rate could be due to a broad range of infections that could be treated using Amoxicillin, including but not limited to chest (bronchitis or pneumonia), blood (septicemia), ears (otitis media), teeth, sinuses (sinusitis), abdomen (intra-abdominal sepsis and peritonitis), heart (endocarditis), kidneys and skin (see (Mousazadeh et al. 2015)).

Benefiting from the experts' ideas involving some managers and specialists from both "Tamin Pharmaceutical Investment Company" (one of the largest production and distribution pharmaceutical companies in Iran) and well-known "Red Crescent Society of Iran", they considered all the 31 provinces in Iran as potential locations for opening LDCs. Among all the provinces, eight of them were selected as candidate locations for opening the MSs, i.e., Tehran, Esfahan, Mashhad, Tabriz, Kermanshah, Karaj, Ghazvin, and Shiraz and finally ten provinces including the candidate provinces for establishing manufacturing centers accompanying two other provinces (i.e. Ahwaz and Mazandaran) for opening MDCs. The geographical representation of candidate locations is represented in Fig. 18.8.

In addition, four possible production capacity levels for MSs and three potential storage capacity levels for MDCs were suggested by experts. They also defined the length of each period as 3 months since the usage rate of Amoxicillin drastically changes in different seasons, which surely affects the production, inventory and flow amounts of that product in different seasonal periods.

The collection of required data was rather burdensome. In detail, in order to estimate the demand data for the period 2014–2017, they first collected the consumption data from 2010 to 2014 and then using the linearized trend of these historical data, the yearly demands for the upcoming 4 years were estimated. On the other hand, according to the experts' knowledge, the consumption rates of Amoxicillin is different in different seasons, i.e. spring (Sp) 11%, summer (Su) 9%, autumn (A) 30% and winter (W) 50%. Benefiting from these data, a seasonal, but national consumption rates were then estimated. Since the available historical data was not divided for each province (demand zones), they estimated the demand of each province based on the fraction of the population of each province to the total population of Iran.

In addition, through consulting with shipping agencies, the transportation cost was calculated for all pairs of 31 provinces through taking two steps, i.e. (1) calculating a unit transportation cost per each 100 km, and (2) multiplying the distance between each pair of provinces with the estimated data from the first step. Also, knowing the volume of each package containing 100 amoxicillin 500 mg capsules as well as an average cost of renting a suitable storage facility with a given capacity, the storage cost per ton of product has been calculated.

It is clear that taking the explained steps will only lead to making a rough estimate. Hence, the authors applied robust possibilistic approach in order to handle epistemic uncertainty in uncertain parameters as well as achieving robust solutions

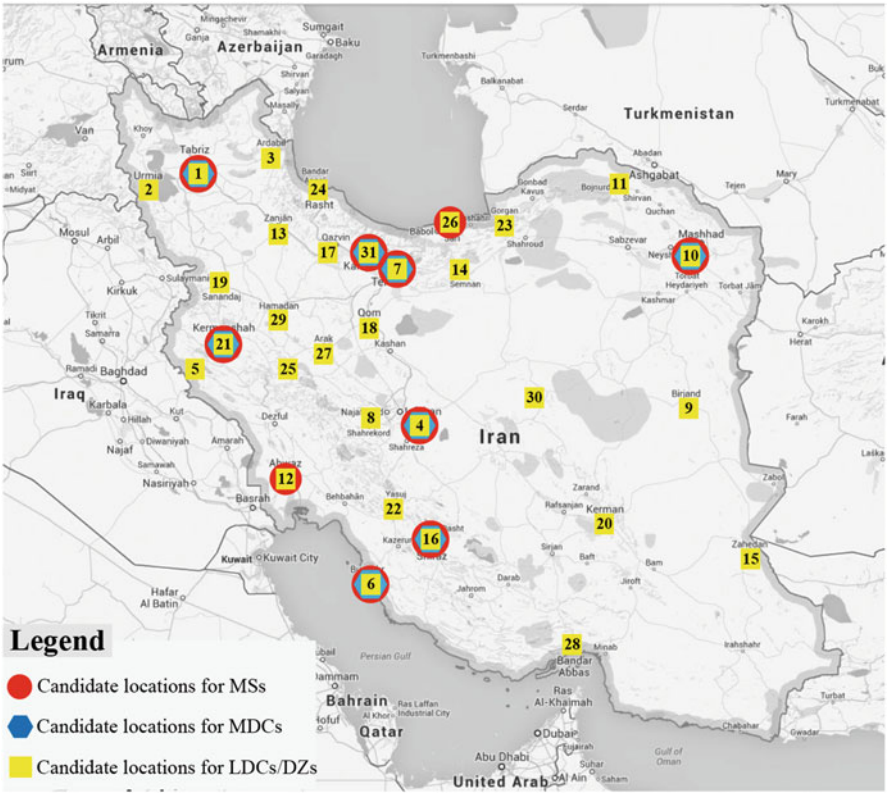

Fig. 18.8 Representation of candidate locations for MSs, MDCs, LDCs and location of DZs

in terms of both feasibility robustness and optimality robustness. In addition, in order to achieve a final compromise solution for the proposed bi-objective model, they utilized the TH approach proposed by Torabi and Hassini (2008).

In the final optimal solution, one MS and one MDC with their least production and storage capacity were opened in Mashhad. In addition, 9 LDCs in Khorasan-e Jonoubi, Khorasan-e Shomali, Zanjan, Semnan, Ghom, Kerman, Lorestan, Markazi and Yazd support MS and MDC facilities in distributing Amoxicillin 500 mg Caps to all 31 provinces/demand zones. Also, the achieved solution leads to the suitable satisfaction degrees, i.e. 0.793 and 0.790 for the first and second objective functions, respectively. A comprehensive sensitivity analysis over the importance weight of each objective function and its effect on optimal design of the national PSCN is also proposed in the chapter.

As could be seen in Fig. 18.9, the optimal solution suggests producing Amoxicillin in all periods (but less in springs) since the model puts an upper bound on production capacity in each period. On the other hand, inventory is being held mostly in two seasons, i.e. summer and autumn in order to cover the high demand for

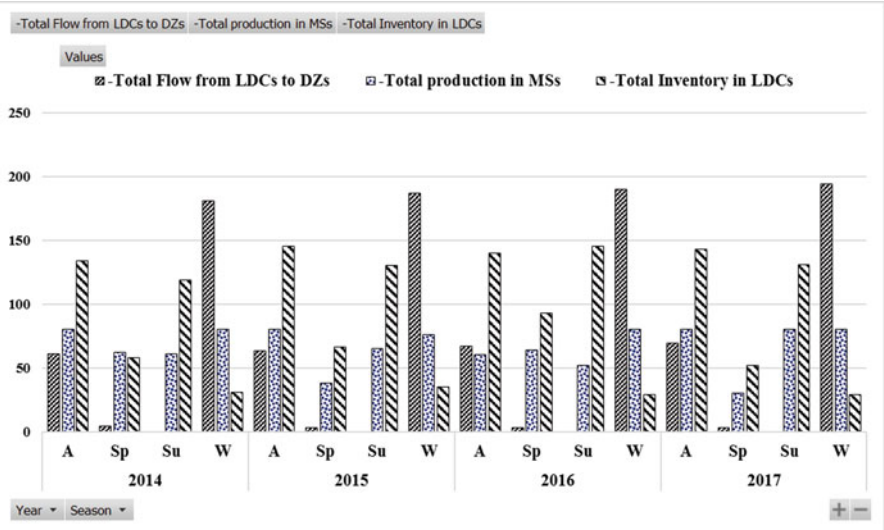

Fig. 18.9 Representation of final compromised solution (Adapted from Mousazadeh et al. (2015))

the drug in winter, the season in which production could not exceed the maximum production capacity. Surprisingly, due to the high holding cost of drug in MDCs, no inventory is held in MDCs at the end of each period. In fact, an established MDC in Mashhad will function only as a national distribution center rather than a large warehouse.

18.5 Future Research Directions

Given the novel concept of PSCs and their special characteristics compared with other SCs, there are many challenges and research gaps in this area, among them we refer to the following ones:

- To the best of our knowledge, environmental and especially the social aspects of PSCs, which beside the economic dimension lead to sustainable frameworks, have not been taken into account in the current literature. So, due the importance of sustainability and the pressures to expend it, an interested research area can be the incorporation of sustainability issues in the PSCs problems. This situation potentially calls for using multi-criteria optimization techniques.
- The concept of reverse logistics and closed-loop SCs, which are focusing on collecting and reusing products at the end of their lifecycles, has received many attentions in the recent years. It provides this opportunity for companies both to obtain economic gains due to recycling of products and so reducing the use of raw material, as well as to comply with environmental legislations (Papageorgiou

2009). This approach has been applied in different industries such as automotive (Schultmann et al. 2006), battery (Kannan et al. 2010), carpet, electrical and electronic equipment (Quariguasi Frota Neto et al. 2010), home appliance (Shih 2001), copier (Krikke et al. 1999), etc. However, limited academic works have been conducted in chemical and particularly pharmaceutical industries.

- The optimization of PSCs under uncertainty have received little attention, in despite of the plenteous methodologies existed in the literature.
- The large size and complexity of real-life problems debilitates the exact methods to reach the solution of the problems, even though they may be equipped with decomposition techniques. Therefore, there is a serious need to develop powerful solution methodologies such as heuristics and meta-heuristics; especially the meta-heuristics.
- Since the integrated continuous production scheme introduced in the first section is a novel concept in the pharmaceutical industry, it brings many challenges to be investigated by researchers.
- Also, the published researches for PSCs have focused on some parts of the PSCs, not all the echelons in an integrated manner. Therefore, considering all echelons of the PSCs simultaneously to be optimized as integrated decision models would be worthwhile.

As a final point, due to the fact that chemical industry establishes the backbone of the pharmaceutical industry and also pursuing the rich literature of this field, in comparison with the young background of the pharmaceutical industry, many ideas can be inspired from the older one to help in the progress of the younger.

References

Amaro A, Barbosa-Póvoa APF (2008) Planning and scheduling of industrial supply chains with reverse flows: a real pharmaceutical case study. Comput Chem Eng 32(11):2606–2625

Antonijevic Z (2014) Optimization of pharmaceutical R&D programs and portfolios. Springer International Publishing, Cham

Bassett MH, Pekny JF, Reklaitis GV (1996) Decomposition techniques for the solution of large-scale scheduling problems. AICHE J 42(12):3373–3387

Birewar DB, Grossmann IE (1989) Incorporating scheduling in the optimal design of multiproduct batch plants. Comput Chem Eng 13(1):141–161

Blau GE, Pekny JF, Varma VA, Bunch PR (2004) Managing a portfolio of interdependent new product candidates in the pharmaceutical industry. J Prod Innov Manag 21(4):227–245

Blomer F, Gunther H-O (2000) LP-based heuristics for scheduling chemical batch processes. Int J Prod Res 38(5):1029–1051

Blömer F, Günther H-O (1998) Scheduling of a multi-product batch process in the chemical industry. Comput Ind 36(3):245–259

Bok J-K, Grossmann IE, Park S (2000) Supply chain optimization in continuous flexible process networks. Ind Eng Chem Res 39(5):1279–1290

Bowman EH (1959) The schedule-sequencing problem. Oper Res 7(5):621–624

Burns L (2002a) The health care value chain: producers, purchasers, and providers. Jossey-Bass, Wharton School Colleagues, San Francisco

Burns L (2002b) The health care value chain: producers, purchasers, and providers. Jossey-Bass, San Francisco
Caillet T (2015) SCM in a pharmaceutical company. In: Supply chain management and advanced planning. Springer, Heidelberg, pp 459–474
Chiou C-C (2008) Transshipment problems in supply chainsystems: review and extensions. INTECH Open Access Publisher, Shanghai
Colvin M, Maravelias CT (2008) A stochastic programming approach for clinical trial planning in new drug development. Comput Chem Eng 32(11):2626–2642
Colvin M, Maravelias CT (2010) Modeling methods and a branch and cut algorithm for pharmaceutical clinical trial planning using stochastic programming. Eur J Oper Res 203(1):205–215
Dimitriadis A, Shah N, Pantelides C (1997) RTN-based rolling horizon algorithms for medium term scheduling of multipurpose plants. Comput Chem Eng 21:S1061–S1066
Dubois D, Fargier H, Fortemps P (2003) Fuzzy scheduling: modelling flexible constraints vs. coping with incomplete knowledge. Eur J Oper Res 147(2):231–252
Eilon S (1957) Scheduling for batch production. Prod Eng J Inst 36(9):549–570
Gatica G, Papageorgiou L, Shah N (2003) Capacity planning under uncertainty for the pharmaceutical industry. Chem Eng Res Des 81(6):665–678
Gatica G, Shah N, Papageorgiou LG (2001) Capacity planning under clinical trials uncertainty for the pharmaceutical industry. Comput Aided Chem Eng 9:865–870
Geletu A, Li P (2014) Recent developments in computational approaches to optimization under uncertainty and application in process systems engineering. Chem Bio Eng Rev 1(4):170–190
Grossmann IE, Quesada I, Raman R, Voudouris VT (1996) Mixed-integer optimization techniques for the design and scheduling of batch processes. Batch processing systems engineering. Springer, pp 451–494
Grunow M, Günther H-O, Yang G (2003) Plant co-ordination in pharmaceutics supply networks. OR Spectr 25(1):109–141
Gupta A, Maranas CD (1999) A hierarchical Lagrangean relaxation procedure for solving midterm planning problems. Ind Eng Chem Res 38(5):1937–1947
Hansen KRN, Grunow M (2015) Planning operations before market launch for balancing time-to-market and risks in pharmaceutical supply chains. Int J Prod Econ 161:129–139
Iyer RR, Grossmann IE (1998) A bilevel decomposition algorithm for long-range planning of process networks. Ind Eng Chem Res 37(2):474–481
Jou C (2005) A genetic algorithm with sub-indexed partitioning genes and its application to production scheduling of parallel machines. Comput Ind Eng 48(1):39–54
Kannan G, Sasikumar P, Devika K (2010) A genetic algorithm approach for solving a closed loop supply chain model: a case of battery recycling. Appl Math Model 34(3):655–670
Karaesmen IZ, Scheller-Wolf A, Deniz B (2011) Managing perishable and aging inventories: review and future research directions. In: Planning production and inventories in the extended enterprise. Springer, New York, pp 393–436
Kelle P, Woosley J, Schneider H (2012) Pharmaceutical supply chain specifics and inventory solutions for a hospital case. Oper Res Health Care 1(2):54–63
Kondili E, Pantelides C, Sargent R (1993) A general algorithm for short-term scheduling of batch operations—I. MILP formulation. Comput Chem Eng 17(2):211–227
Krikke H, van Harten A, Schuur P (1999) Business case Oce: reverse logistic network re-design for copiers. OR-Spektrum 21(3):381–409
Ku HM, Karimi I (1991) An evaluation of simulated annealing for batch process scheduling. Ind Eng Chem Res 30(1):163–169
Laínez JM, Schaefer E, Reklaitis GV (2012) Challenges and opportunities in enterprise-wide optimization in the pharmaceutical industry. Comput Chem Eng 47:19–28
Levis AA, Papageorgiou LG (2004) A hierarchical solution approach for multi-site capacity planning under uncertainty in the pharmaceutical industry. Comput Chem Eng 28(5):707–725
Maravelias CT, Sung C (2009) Integration of production planning and scheduling: overview, challenges and opportunities. Comput Chem Eng 33(12):1919–1930

Masoumi AH, Yu M, Nagurney A (2012) A supply chain generalized network oligopoly model for pharmaceuticals under brand differentiation and perishability. Trans Res E: Logist Transp Rev 48(4):762–780

Mauderli A, Rippin D (1979) Production planning and scheduling for multi-purpose batch chemical plants. Comput Chem Eng 3(1):199–206

Méndez CA, Cerdá J, Grossmann IE, Harjunkoski I, Fahl M (2006) State-of-the-art review of optimization methods for short-term scheduling of batch processes. Comput Chem Eng 30(6):913–946

Mousazadeh M, Torabi S, Zahiri B (2015) A robust possibilistic programming approach for pharmaceutical supply chain network design. Comput Chem Eng 82:115–128

Mousazadeh M, Torabi SA, Pishvaee MS (2014) Green and reverse logistics management under fuzziness. In: Supply chain management under fuzziness. Springer, Heidelberg, pp 607–637

Mula J, Poler R, Garcia J (2006) MRP with flexible constraints: a fuzzy mathematical programming approach. Fuzzy Sets Syst 157(1):74–97

Narayana SA, Pati RK, Vrat P (2012) Research on management issues in the pharmaceutical industry: a literature review. Int J Pharm Healthc Mark 6(4):351–375

Naresh S (2012) Planning and scheduling in pharmaceutical supply chains. Engineering Department of Chemical and Biomolecular Engineering. National University of Singapore

Oh H-C, Karimi I (2004) Regulatory factors and capacity-expansion planning in global chemical supply chains. Ind Eng Chem Res 43(13):3364–3380

Oh HC, Karimi I (2006) Global multiproduct production–distribution planning with duty drawbacks. AICHE J 52(2):595–610

Orcun S, Altinel I, Hortaçsu Ö (2001) General continuous time models for production planning and scheduling of batch processing plants: mixed integer linear program formulations and computational issues. Comput Chem Eng 25(2):371–389

Papageorgiou LG (2009) Supply chain optimisation for the process industries: advances and opportunities. Comput Chem Eng 33(12):1931–1938

Papageorgiou LG, Pantelides CC (1996) Optimal campaign planning/scheduling of multipurpose batch/semicontinuous plants. 1. Mathematical formulation. Ind Eng Chem Res 35(2):488–509

Papageorgiou LG, Rotstein GE, Shah N (2001) Strategic supply chain optimization for the pharmaceutical industries. Ind Eng Chem Res 40(1):275–286

Pinto JM, Grossmann IE (1995) A continuous time mixed integer linear programming model for short term scheduling of multistage batch plants. Ind Eng Chem Res 34(9):3037–3051

Pishvaee M, Razmi J, Torabi SA (2012) Robust possibilistic programming for socially responsible supply chain network design: a new approach. Fuzzy Sets Syst 206:1–20

Pishvaee MS, Khalaf MF (2016) Novel robust fuzzy mathematical programming methods. Appl Math Model 40(1):407–418

Pritsker AAB, Waiters LJ, Wolfe PM (1969) Multiproject scheduling with limited resources: a zero-one programming approach. Manag Sci 16(1):93–108

Quariguasi Frota Neto J, Walther G, Bloemhof J, Van Nunen J, Spengler T (2010) From closed-loop to sustainable supply chains: the WEEE case. Int J Prod Res 48(15):4463–4481

Rogers MJ, Gupta A, Maranas CD (2002) Real options based analysis of optimal pharmaceutical research and development portfolios. Ind Eng Chem Res 41(25):6607–6620

Rotstein G, Papageorgiou L, Shah N, Murphy D, Mustafa R (1999) A product portfolio approach in the pharmaceutical industry. Comput Chem Eng 23:S883–S886

Sahinidis NV (2004) Optimization under uncertainty: state-of-the-art and opportunities. Comput Chem Eng 28(6):971–983

Saur-Amaral I, Borges Gouveia JJ (2007) Uncertainty in drug development: insights from a Portuguese firm. Int J Technol Intell Plan 3(4):355–375

Schaber SD, Gerogiorgis DI, Ramachandran R, Evans JM, Barton PI, Trout BL (2011) Economic analysis of integrated continuous and batch pharmaceutical manufacturing: a case study. Ind Eng Chem Res 50(17):10083–10092

Schultmann F, Zumkeller M, Rentz O (2006) Modeling reverse logistic tasks within closed-loop supply chains: an example from the automotive industry. Eur J Oper Res 171(3):1033–1050

Shah N (2004) Pharmaceutical supply chains: key issues and strategies for optimisation. Comput Chem Eng 28(6):929–941
Shah N, Pantelides C, Sargent R (1993) A general algorithm for short-term scheduling of batch operations—II. Computational issues. Comput Chem Eng 17(2):229–244
Shah N, Pantelides CC (1991) Optimal long-term campaign planning and design of batch operations. Ind Eng Chem Res 30(10):2308–2321
Shelokar P, Jayaraman VK, Kulkarni BD (2004) An ant colony classifier system: application to some process engineering problems. Comput Chem Eng 28(9):1577–1584
Shih L-H (2001) Reverse logistics system planning for recycling electrical appliances and computers in Taiwan. Resour Conserv Recycl 32(1):55–72
Sousa RT, Liu S, Papageorgiou LG, Shah N (2011) Global supply chain planning for pharmaceuticals. Chem Eng Res Des 89(11):2396–2409
Sousa RT, Shah N, Papageorgiou LG (2005) Global supply chain network optimisation for pharmaceuticals. Comput Aided Chem Eng 20:1189–1194
Sundaramoorthy A, Evans JM, Barton PI (2012a) Capacity planning under clinical trials uncertainty in continuous pharmaceutical manufacturing, 1: mathematical framework. Ind Eng Chem Res 51(42):13692–13702
Sundaramoorthy A, Karimi I (2004) Planning in pharmaceutical supply chains with outsourcing and new product introductions. Ind Eng Chem Res 43(26):8293–8306
Sundaramoorthy A, Li X, Evans JM, Barton PI (2012b) Capacity planning under clinical trials uncertainty in continuous pharmaceutical manufacturing, 2: solution method. Ind Eng Chem Res 51(42):13703–13711
Susarla N, Karimi IA (2012) Integrated supply chain planning for multinational pharmaceutical enterprises. Comput Chem Eng 42:168–177
Tollman P, Morieux Y, Murphy J, Schulze U (2011) Can R&D be fixed? Lessons from biopharma outliers. The Boston Consulting Group, Boston
Torabi SA, Hassini E (2008) An interactive possibilistic programming approach for multiple objective supply chain master planning. Fuzzy Sets Syst 159(2):193–214
Traple MAL, Saviano AM, Francisco FL, Lourenço FR (2014) Measurement uncertainty in pharmaceutical analysis and its application. J Pharmaceut Anal 4(1):1–5
Tsang K, Samsatli N, Shah N (2007) Capacity investment planning for multiple vaccines under uncertainty: 1: capacity planning. Food Bioprod Process 85(2):120–128
Üster H, Easwaran G, Akçali E, Cetinkaya S (2007) Benders decomposition with alternative multiple cuts for a multi-product closed-loop supply chain network design model. Naval Res Logist (NRL) 54(8):890–907
Uthayakumar R, Priyan S (2013) Pharmaceutical supply chain and inventory management strategies: Optimization for a pharmaceutical company and a hospital. Oper Res Health Care 2(3):52–64
Varma VA, Pekny JF, Blau GE, Reklaitis GV (2008) A framework for addressing stochastic and combinatorial aspects of scheduling and resource allocation in pharmaceutical R&D pipelines. Comput Chem Eng 32(4):1000–1015
Vila-Parrish AR, Ivy JS, King RE, Abel SR (2012) Patient-based pharmaceutical inventory management: a two-stage inventory and production model for perishable products with Markovian demand. Health Syst 1(1):69–83
Wellons MC, Reklaitis GV (1991) Scheduling of multipurpose batch chemical plants. 2. Multiple-product campaign formation and production planning. Ind Eng Chem Res 30(4):688–705
Wilkinson S, Shah N, Pantelides C (1995) Aggregate modelling of multipurpose plant operation. Comput Chem Eng 19:583–588
Yi G, Reklaitis GV (2007) Optimal design of batch-storage network considering exchange rates and taxes. AICHE J 53(5):1211–1231

Chapter 19
A Categorical DEA Framework for Evaluating Medical Tourism Efficiency of "Top Destinations"

Melis Almula Karadayi and Mine Isik

19.1 Introduction

One of the megatrends of the twenty-first century will continue to be increased mobility, enabled by tourism. The concept of tourism is not confined to leisure tourism, but includes adventure tourism, pilgrimage tourism, health tourism, and eco-tourism. Among different types of tourism, medical tourism is becoming more and more popular. For many countries, the concept of medical tourism appears to be challenging with enormous potential including a direct and indirect impact on gross domestic product and foreign exchange earnings.

According to World Health Organization (WHO), there exists no agreed definition of medical tourism. As Yu and Ko (2012) describe medical tourism as a "combination of medical services and the tourism industry", it is beneficial first to pay attention to how World Tourism Organization (WTO) describes tourism.

WTO describes it as follows: "Tourism comprises the activities of persons traveling to and staying in places outside their usual environment for not more than one year for leisure, business and other purposes not related to the exercise of an activity remunerated from within the places visited." The term 'other purposes' includes "traveling across international borders to receive some form of medical treatment" as WHO describes medical tourism. Medical Tourism Association (2016) tries to construct a more comprehensive definition of medical tourism and

M.A. Karadayi
Industrial Engineering Department, Medipol University, Beykoz, Istanbul, 34810, Turkey
e-mail: melisalmula.sye@gmail.com

M. Isik (✉)
Industrial Engineering Department, Bogazici University, Bebek, Istanbul, 34342, Turkey
e-mail: makaradayi@medipol.edu.tr

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_19

lists the main arguments of travelling abroad to receive medical care. Better access to higher quality medical operations and low cost are considered as the main drivers.

Thanks to technological advances, more affordable transportation and positive change in regulatory framework as well as increase in the number of health professionals, people tend to get medical care from foreign countries more often. These push factors have a multidimensional structure. Both social and financial drives make domestic medical care disadvantageous.

Long waiting lists due to the global growth in the number of people that require medical care, relatively higher costs of elective treatment in developed countries, and some strict governmental procedures (for some specific surgical operations including organ transplantation, fertility treatment etc.) encourage people to seek medical care abroad since this situation makes it hard to get the necessary treatment at the right time. It also contributed to the movement of patients across countries and boosted medical tourism in the past decade (Lee 2010).

Apart from "push factors" that have a negative effect on the eligibility of domestic medical care, there exists "pull factors" which make foreign countries more attractive.

On the other hand, medical care is quite expensive in advanced countries; and developing countries have a potential to propose the same quality of health care services at a relatively low cost especially in the past 15 years (Beladi et al. 2015). In addition to the money-saving opportunity, the availability of some particular treatment type in another country, higher quality of care, timeliness (in contrast to long waiting lists in developed countries) could also be considered as pull factors (Glinos et al. 2010).

In the literature, marketing and governmental actions that promote medical tourism is the main point of concern. Additionally, instead of dealing with the whole medical tourism sector, most of the researchers focus on a subset of medical tourism (hair transplantation, esthetic or dental surgery operations. Moreover, the research area of medical tourism performance evaluation of countries currently lacks sufficient exploration. Medical tourism is a great opportunity that could generate foreign currency (Pan and Chen 2014).

The purpose of this study is to present a DEA framework for evaluating the medical tourism performance of top destination countries. In this regard, the objective is to define appropriate efficiency scores that have the power to represent selected countries' medical tourism performance.

As a result, this study will contribute to the existing literature by presenting a model that is able to reveal a road map that explores the strengths and weaknesses that belong to the top medical tourism destinations with respect to the best practice (Decision making unit that has the highest efficiency value). Furthermore, the results of the study would also provide information to benefit the government of the top medical tourism destinations for policy evaluation of medical tourism in a specific country.

The rest of the chapter is structured as follows. The following section outlines a review of medical tourism. The basics of DEA and categorical DEA models are delineated in Sect. 3. Application of the proposed methodology for evaluating medical tourism efficiency of top destinations is presented in Sect. 4. Conclusions and future research directions are presented in final section.

19.2 Literature Review

This section outlines a review of studies about medical tourism and also the application of DEA to important problems in health care.

The first group of studies include medical tourism studies. Medical tourism is the major player in shaping the globalization of health services, considering positively changing conditions in technology, economy, governmental and international relations (Mohamad et al. 2012). Under these more eligible conditions, medical tourists can be defined as patients that travel internationally in order to get medical treatments including organ transplants, cosmetic, and dentistry operations, etc. (Chuang et al. 2014). As the awareness about various positive drivers and motivation factors increased, the range of medical tourism continues to increase apart from the above-mentioned options. As Lunt et al. (2016) mention, the motivation of traveling overseas for medical care is complex. To solve this comprehensive motivation structure, in the literature, pull (familiarity, quality, etc.) and push factors (treatment delays, long waiting lists, cost etc.) are commonly referenced (Glinos et al. 2010). Researchers aspire to reveal the demand factors of medical tourists and the effect of supply characteristics that belong to destination countries.

Rozee and Rochebrochard (2013) explore one of the push factors (long waiting lists) for the case of French patients. Although they are eligible for treatment, on the occasion of oocyte shortage in France, patients prefer to travel abroad for care.

The study of Van Hoof et al. (2015) also analyze the negative effect of legal restrictions on domestic treatments and explain how it can push patients to get treatment in another country.

Veerasoontorn and Beise-Zee (2010) prepared a survey with 27 respondents (medical tourists traveling to Thailand) in order to identify push and pull factors. Results expose that worsening service conditions and high costs are the negative factors that prevent patients from domestic medical care. On the other hand, innovation, efficiency and service quality are motivators. Crooks et al. (2011) also acknowledge that the quality of the medical treatment is a vital fact that shapes patients' decision on hospital selection.

The Medical Tourism Survey prepared by Medical Tourism Association (2013) reveals the opinions of medical tourists on health care facility selection. Apart from health care service quality as a whole, qualification, and availability of the physicians (doctors/dentist) are believed to be essential factors by respondents.

Furthermore, Alsharif et al. (2010) investigated patients' (that traveling to Jordan, China, India, the United Arab Emirates aiming to get medical treatment) motivation drivers and discovered that the "physician" is one of the important decision-making criteria on health care treatment facility.

Apart from qualifications of "medical services", since patients are able to combine their medical operations with sightseeing or touristic attractions during the operation and recovery period, countries' natural and historical value has a direct impact on medical tourists' choice of destination. This relationship between health care and international tourism potential drew researchers' interest. In the study of

Lee (2010), the bounds testing approach and Granger causality test are applied to explore the dynamic interrelation between medical treatment and international tourism both for the short and long run for the Singapore case. The results reveal that there exists a unidirectional interaction in the long-run. Yu and Ko (2012) also utilize an exploratory factor analysis to exhibit the relationship between tourism activities and health care services and the attractiveness of the country (both natural and historical); and found it influential.

Additionally, Pan and Chen's research provides corroborative results that there is a significant relationship between tourism and attractiveness of a medical care destination (2014).

Apart from cost, service quality and touristic attractions, researchers claim that political environment and stability constitute another fundamental issue on medical tourism (Fetscherin and Stephano 2016; Smith et al. 2011).

All of the above-mentioned studies try to reveal the main drivers to give a reliable answer to the question of "Why people travel to get medical treatment?" In spite of its importance, there is not an adequate level of research on the interrelationship between medical care and tourism in any of the existing few studies; only a couple are conducted that utilize operations research methods. One of them is constructed by Bies and Zacharia (2007). They establish an Analytic Network Process for Decision-Making) model to reveal whether government sponsored programs are preferable compared to self-selected medical tourism. The model consists of SWOT elements, and the objective is to find the weights of these elements to evaluate the alternative medical tourism destinations.

Numerous nations have significantly benefited from medical tourism. On the other hand, there is a lack of attention on the country performance that reveals the steps that should be taken to increase the number of medical tourists.

Furthermore, because primary drivers of patients' preferences on medical tourism are complicated, this study first presents the factors selected for model construction and gives relative performance scores that belong to top medical tourism destinations.

In the second group, DEA applications in health care were analyzed. In the literature, the first study of DEA in the health care field is conducted by Wilson and Jadlow (1982). Zere et al. (2006) employed the constant returns to scale (CRS) model to analyze hospital performance performance of 26 district hospitals in Namibia. Hajialiafzali et al. (2007) evaluated efficiency of Iranian social security organization hospitals. Nayar and Ozcan (2008) investigated the efficiency of hospitals in Virginia using input-oriented CRS models. Mark et al. (2009) investigated the technical efficiency of 226 healthcare units using variable returns to scale (VRS) model. Dash et al. (2010) employed input-oriented approach with VRS to evaluate 29 district hospitals of Tamil Nadu. Ketabi (2011) applied the input-oriented and VRS model to assess the efficiency of 23 cardiac care units in Iran. Hu et al. (2012) examined the performance of hospitals in 30 province-level units in China using the BCC model. Gautam et al. (2013) employed healthcare efficiency of hospitals in Missouri using an input-oriented VRS model. De Nicola et al. (2014) suggested a DEA approach in order to analyze the healthcare efficiency of regions in Italy.

Lately, Prakash and Annapoorni (2015) examined the healthcare performance of public hospitals using an output-oriented BCC model in Tamil Nadu for the year 2012 and 2013.

19.3 Data Envelopment Analysis

DEA is method that utilizes the main assumptions of a linear programming to measure the efficiency of decision making units (DMUs) relative to the DMU that has the highest efficiency value. It can accommodate multiple input and output data. DEA gives relative efficiency of each DMU by calculating the ratio of its total weighted output to its total weighted input. The objective function of the target-DMU is formed by this ratio. The normalizing constraints are added to the model to ensure that input-output ratio that belong to the DMU can be unity. The mathematical model is given below:

$$\max\ E_{j_0} = \frac{\sum_r u_r y_{rj_0}}{\sum_i v_i x_{ij_0}}$$

Subject to

$$\frac{\sum_r u_r y_{rj}}{\sum_i v_i x_{ij}} \leq 1, \quad j = 1, \cdots, n$$

$$u_r, v_i \geq \varepsilon > 0 \quad r = 1, \cdots s; i = 1, \cdots, m$$

Where
E_{jo} = *efficiency value of* j_0
u_r = *weight of output r*
v_i = *weight of input i*
x_{ij} = *amount of input i used by the DMU j*
y_{ij} = *amount of output r produced by the DMU j*
ε = *infinitesimal positive number*
n = *number of DMUs*
m = *number of inputs*
s = *number of outputs*

The efficiency score of a DMU can vary between 0 and 1. If a DMU takes a value of 1, it means that it is relatively efficient. On the other hand, if it takes a value less than 1, it indicates that it is an inefficient DMU. DEA offers input and output orientations to efficiency analysis. In DEA method that is structured as an input oriented, if a DMU is inefficient, by proportionally reducing its input values, while the proportions of outputs are held constant. On the other hand, in an output-oriented DEA, an inefficient DMU is able to reach an efficient score by increasing the output values, while the proportion of inputs remain the same.

The standard DEA methods can evaluate the relative efficiencies of DMUs without knowing production functions, only by considering input and output values (Zhou et al. 2012). In a case where a set of establishments are at different levels, in order to make a fair evaluation of each DMU, a DMU from any category is allowed to be compared only to the DMUs which are in the same or less-advantageous categories. In order to deal with this situation, categorical DEA models are preferred (Cook and Seiford 2009).

19.3.1 Data Envelopment Analysis with Categorical Data

Banker and Morey (1986) suggested the first categorical method in DEA. If we have categorical variables in an efficiency analysis, a methodology is required for ensuring that the composite reference members be constituted from DMUs that belong to the same category or possibility clustering in the same category which is assumed to be operating in a more difficult or objectionable situation (Banker and Morey 1986). In this context, DMUs are divided into sub-groups based on the categorical variables. If there are L (1,2,...L) levels of a categorical input variable, these L values divide the DMU into subgroups.

When a DMU is evaluated based on the surface which is formed by the DMUs that are in its and the previous categories, a $DMU \in D_K$ which will be evaluated with respect to the DMUs in $U_{k=1}^{k} D_K$ considers $k \in \{1, \ldots, L\}$. All $I \in D_1$ DMUs are evaluated according to the DMUs in D_1, $I \in D_2$ DMUs are evaluated with respect to the DMUs in D_1, D2, $I \in D_c$ DMUs are evaluated with respect to the DMUs in$U_{k=1}^{c} D_K$, and $I \in D_L$ DMUs are evaluated with respect to the DMUs in $U_{k=1}^{L} D_K$ (Banker and Morey 1986; Taskopru 2014).

19.3.1.1 CCR Model with Categorical Variables

Charnes et al. (1978) model with categorical variables relies on the assumption of CRS. The following model is the input-oriented CCR with categorical variables:

$$min\theta_k$$

$$\text{Subject to} \quad \begin{array}{l} \sum_{j \in U_{f=1}^{k} D_f}^{n} \lambda_j X_{ij} \leq \theta_k X_{ik} \\ \sum_{j \in U_{f=1}^{k} D_f}^{n} \lambda_j Y_{rj} \leq Y_{rk} \\ \sum_{j \in U_{f=1}^{k} D_f}^{n} \lambda_j = 1 \\ \lambda_j \geq 0 \; and \; j \in U_{f=1}^{L} D_f \end{array}$$

Where
$\theta_k = efficiency\ value\ of\ DMU\ k$
$Y_{rk} = r^{th}\ output\ value\ of\ DMU\ k$
$Y_{rj} = r^{th}\ output\ value\ of\ DMU\ j$
$X_{ik} = i^{th}\ input\ value\ of\ DMU\ k$
$Xij = i^{th}\ input\ value\ of\ DMU\ j$
$\lambda_j = weight\ of\ DMU\ j$
$n = numbers\ of\ DMUs$
$L = number\ of\ categories$

The following mathematical model is a version of dual classical output-oriented CCR model with categorical variables:

$$max\varphi_k$$

$$\text{Subject to} \quad \begin{array}{l} \sum_{j\in U_{f=1}^{k}}^{n} {}_{D_f}\eta_j X_{ij} \leq X_{ik} \\ \varphi_k Y_{rk} - \sum_{j\in U_{f=1}^{k}}^{n} {}_{D_f}\eta_j Y_{rj} \leq 0 \\ \eta_j \geq 0 \ \ and \ \ j \in U_{f=1}^{L} D_f \end{array}$$

Where
φ_k = *efficiency value of DMU k*
η_j = *weight of DMU j*

19.4 Application of the Proposed DEA Framework for Evaluating Medical Tourism Efficiency of Top Destination Countries

Factors such as affordability, quality of care provided, alternative/innovative therapy options and waiting time for care are considered as the main drivers of the medical tourism attractiveness. According to the level of these factors, top destinations for medical tourism are categorized as "key destinations", "emerging destinations" and "other destinations" depending on the country's medical competencies and capabilities (Grail Research 2009). Since the accurate measurement of countries in terms of medical tourism efficiency requires a hierarchical category. To deal with this issue, this study proposes a categorical tourism performance of destinations which are Mexico, Brazil, India, Thailand, Singapore, Malaysia, the United Arab Emirates, South Korea, Taiwan, Chile, Cuba, Panama, Argentina, South Africa, Jordan, Israel and Turkey.

Medical tourism could be analyzed from different perspectives or by investigating a variety of forces that play in a field that is establishing itself (Mainil 2012). Benefiting from the literature and considering the availability of data, five inputs, namely "categorical variable", "health expenditure", "global peace index", "number of physicians", "tourism index" with only one output parameter "number of medical tourists" are taken into account.

It is worth mentioning that this study was purely based on data collected from various reputable sources such as the World Tourism Organization, the World Trade Organization, World Bank, the World Economic Forum and other relevant websites.

The constructed model provides the efficiency index that aims to reveal the value of performance that belongs to each DMU. As a result, for each DMU factor, base improvement opportunities make the country more attractive for medical tourists. Each factor is added into the DEA for a specific purpose. Tourism index stands for the status of the tourism sector of the considered country. Since in the literature review, previous studies reveal that there a strong relationship between "tourist attractions" and "medical tourism", it is substantive to include a tourism index into the constructed model (Yu and Ko 2012; Pan and Chen 2014).

In order to track the effect of the political status of the country, Global Peace Index (GPI) that aims to measure the relative position of the countries' peacefulness constituted by the Institute for Economics and Peace[1] (2015) is included in the DEA model. GPI depicts the nations' level of peacefulness. The index contains three principal parameters; the safety and security level in public, the degree of conflict (both national and international) and the level of militarization.

The Medical Tourism Survey prepared by Medical Tourism Association (2013) can be considered one of the important data sources since there is not a comprehensive survey of international medical tourists' preferences. Since the availability of the physicians (doctors/dentist) appeared to be the most important factor by respondents, number of physicians (data is published by the World Bank) is added into the model.

Health Expenditure % of GDP data is gathered from the World Bank database. The definition of health expenditure is given as follows: It "consists of recurrent and capital spending from government (central and local) budgets, external borrowings and grants (including donations from international agencies and nongovernmental organizations), and social (or compulsory) health funds". In this study, it is assumed that more spending implies providing better health care services due to high budget that is reserved to the investments on health sector.

The categorical variable is set according to the classification presented by Grail Research Institute (which is one of the leading institutions that release reports on medical tourism). Three different categories on significant players in medical tourism are set as "key destinations, emerging destinations and other" considering the number of JCI Accreditation, cost range and popular treatment options (Grail Research 2009).

As an output, the number of medical tourists is taken. It is assumed that the variables used in the efficiency model presented in Table 19.1 have a direct effect on the number of medical tourists. Reciprocal of GPI and tourism index values are considered in the evaluation model. Furthermore, the values of each parameter that belongs to the countries added into the DEA are given respectively in Table 19.2. 2013 is selected as the year of the model due to data availability.

Table 19.1 Description of the variables

	Variable	Description
	Categorical variable	Key, emerging and other medical tourism destinations
Inputs	Health expenditure	Total health expenditure as a percentage of gross domestic product
	Global peace index	Index that measures the peacefulness
	Physicians	Density of physicians per 1000 population
	Tourism index	Index that measures the travel and tourism competitiveness
Outputs	Medical tourists	Total number of medical tourists

[1] An independent think-tank organization.

Table 19.2 Data used in the DEA

DMU	Health expenditure	Global peace index	Physicians	Tourism index	Medical tourists	Category
Cuba	8.02	0.520	6.723	0.224	780,000	3
Panama	11.19	0.528	1.650	0.220	447,660	3
Argentina	4.56	0.524	3.859	0.240	1,416,420	3
South Arica	6.41	0.436	0.776	0.242	500,000	3
Poland	3.73	0.654	2.219	0.224	300,000	3
Hungary	7.63	0.658	3.080	0.222	87,000	3
Jordan	4.87	0.538	2.558	0.239	250,000	3
Israel	7.26	0.366	3.344	0.230	30,000	3
Turkey	9.63	0.410	1.711	0.225	300,000	3
New Zeland	5.43	0.808	2.735	0.193	709,830	3
United Arab Emirates	3.20	0.596	2.533	0.206	135,000	2
South Korea	7.17	0.549	2.143	0.204	399,000	2
Taiwan	9.44	0.650	1.770	0.212	200,000	2
Vietnam	11.03	0.564	1.190	0.253	2,044,440	2
Chile	5.58	0.629	1.024	0.233	965,520	2
Mexico	6.24	0.411	2.095	0.224	1,000,000	1
Costa Rica	8.10	0.570	1.113	0.225	50,000	1
Brazil	3.80	0.488	1.891	0.229	180,000	1
India	5.88	0.389	0.702	0.243	236,898	1
Thailand	4.15	0.421	0.393	0.223	1,800,000	1
Singapore	3.20	0.695	1.950	0.191	610,000	1
Malaysia	7.28	0.635	1.198	0.213	770,134	1
Philippines	4.52	0.421	1.153	0.254	1,263,870	1

Category 1 stands for key destinations. Category 2 stands for emerging destinations and the lowest level, Category 3 stands for other destinations. In this way, medical tourism efficiency of each country could be evaluated only between its own category and the lower categories. CRS assumption is employed. The main argument of selecting the CRS model for evaluating the efficiency scores to examine the input and output correspondence in the absence of any scale or congestion effects (Gok and Sezen 2013). In addition, since we aim to produce the highest possible output levels for a given input usage for each country, we used the output-oriented categorical DEA model. The detailed representation of the proposed framework given in Fig. 19.1 is given below.

The computed technical efficiency scores of top destinations for medical tourism are represented in Fig. 19.2.

The results reveal that a majority of countries are inefficient in terms of medical tourism. The mean technical efficiency of the medical tourism sector is 0.348. "Thailand" and "Vietnam" are determined as the best performing countries according to relative efficiency scores in terms of medical tourism. "Argentina" is ranked as the second country in terms of medical tourism with the efficiency score 0.823. It is followed by the "Philippines," Mexico", and "Chile" with respective efficiency scores: 0.799, 0.569 and 0.529. The lowest value of efficiency is scored by "Israel".

After obtaining inefficient destinations in terms of medical tourism. Potential improvement percentage (PI %) of each input are determined. PI suggestions for each input for inefficient countries in terms of medical tourism are provided in Table 19.3.

Results reveal that PI (%) for tourism index is 0 for all countries since they are top destinations for medical tourism. Besides, there are available PI (%) suggestions for health expenditures, GPI and number of physicians of each inefficient destination.

Negative values for PI (%) correspond to decrease in the actual input values in order to be an efficient country by producing same amount of output.

Studies also reveal that Thailand deserves its efficiency score. Thailand is considered to be the global leader for medical tourism (Medretreat 2016). Thousands of people have been travelling to Thailand due to its world class treatment options with affordable prices. The underlying cause can be attributable to the governmental efforts. A new medical tourism strategy has been established by Thailand; and it is targeted to reach $6 billion medical tourism revenues annually by 2017 (Beladi et al. 2015). Furthermore, the government of Thailand is not only content with money allocation, but also establishes agencies (multi-agency government-industry partnership is also promoted) to enhance their popularity as a medical tourism destination (Lunt et al. 2011; Pocock and Phua 2011).

Vietnam also shows the highest performance with the efficiency score equal to 1. This country establishes a brand name in medical tourism owing to its cut-price medical treatment options by well-qualified physicians. Additionally, Vietnam has much to offer in terms of sightseeing and cultural attractions.

Apart from two popular medical tourism destinations, Argentina gets the third highest efficiency score. If we try to identify the root causes, specialization in

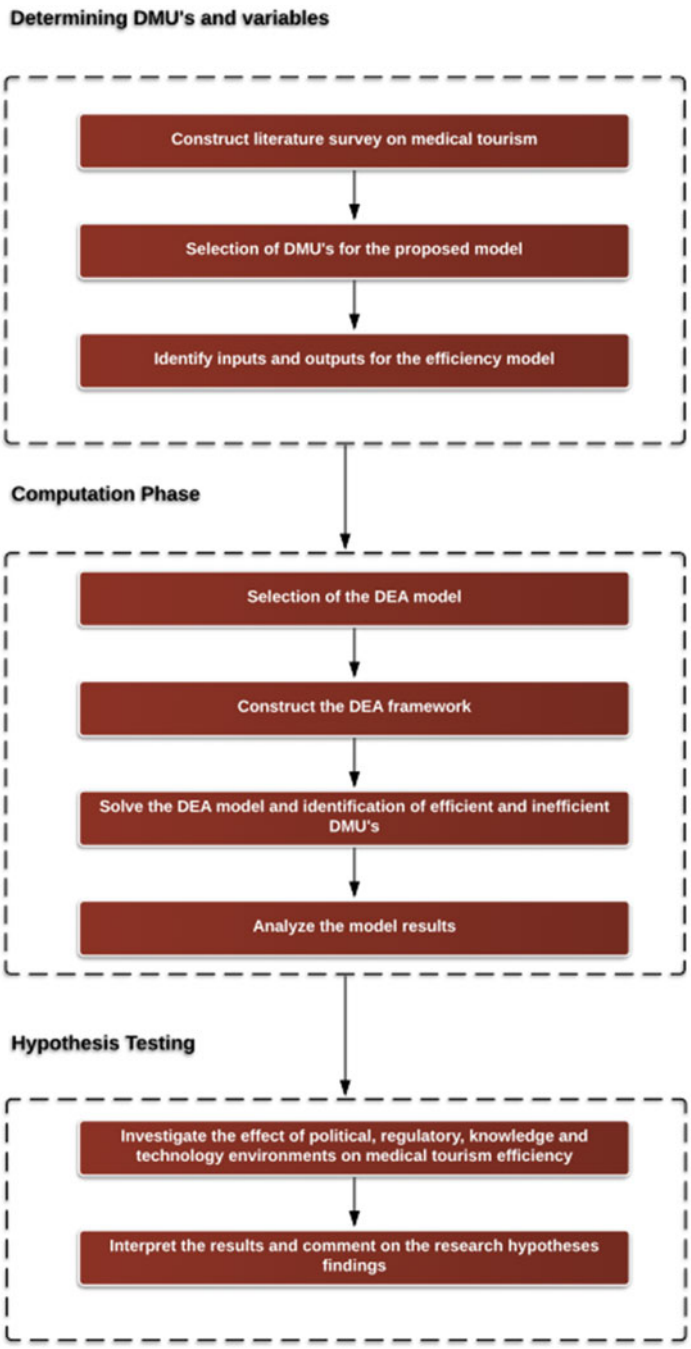

Fig. 19.1 Representation of the proposed framework

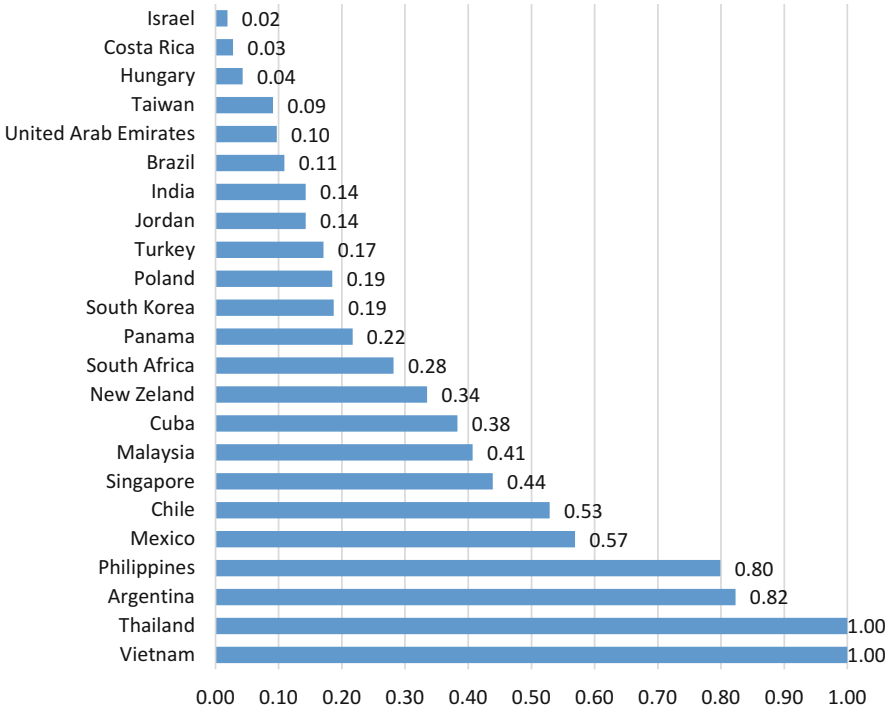

Fig. 19.2 Efficiency scores

cosmetic surgery, bariatric surgery, and breast implants appear to be the main ones. Argentina also aims to develop its own brand, so it marketed itself as a unique medical tourism destination with its natural beauty and top tier cosmetic surgery (Viladrich and Faust 2014).

Singapore and the Philippines are also two of the major players in the medical tourism market. They are known for cardiovascular, neurological surgery and stem cell therapy operations. The government of these countries is evidently promoting medical tourism by allocating their budgets on marketing activities.

The results reveal that Asian nations dominated medical tourism sector by its services. Lack of waiting times resulted from the availability of physicians, rapid developments in surgery operations due to the allocation of a considerable amount of governmental budget on health care sector and natural beauty accompanying low costs are the main drivers as Yu and Ko (2012) assert in their study.

Table 19.3 Potential Improvement suggestions for inefficient medical tourism destinations

	Input	Actual	Target	PI (%)
Cuba	Health Exp.	8.02	7.96	−0.01
	GPI	0.520	0.520	0.00
	Physicians	6.72	0.83	−0.88
	Tourism index	0.22	0.22	0.0
Panama	Health Exp.	11.19	8.08	−0.28
	GPI	0.528	0.528	0.00
	Physicians	1.65	0.84	−0.49
	Tourism index	0.22	0.22	0.0
Argentina	Health Exp.	4.56	4.56	0.00
	GPI	0.524	0.410	−0.22
	Physicians	3.86	0.45	−0.88
	Tourism index	0.24	0.24	0.0
South Africa	Health Exp.	6.41	5.66	−0.12
	GPI	0.436	0.436	0.00
	Physicians	0.78	0.57	−0.26
	Tourism index	0.24	0.24	0.0
Poland	Health Exp.	3.73	3.73	0.0
	GPI	0.654	0.378	−0.42
	Physicians	2.22	0.35	−0.84
	Tourism index	0.22	0.22	0.0
Hungary	Health Exp.	7.63	7.63	0.0
	GPI	0.658	0.514	−0.22
	Physicians	3.08	0.79	−0.74
	Tourism index	0.22	0.22	0.0
Jordan	Health Exp.	4.87	4.87	0.0
	GPI	0.538	0.419	−0.2
	Physicians	2.56	0.48	−0.8
	Tourism index	0.24	0.24	0.0
Israel	Health Exp.	7.26	3.61	−0.5
	GPI	0.366	0.366	0.0
	Physicians	3.34	0.3	−0.9
	Tourism index	0.23	0.23	0.0
Turkey	Health Exp.	9.63	4.05	−0.6
	GPI	0.410	0.410	0.0
	Physicians	1.71	0.38	−0.8
	Tourism index	0.23	0.23	0.0
New Zealand	Health Exp.	5.43	5.43	0.0
	GPI	0.808	0.503	−0.4
	Physicians	2.74	0.53	−0.8
	Tourism index	0.19	0.19	0.0

(continued)

Table 19.3 (continued)

	Input	Actual	Target	PI (%)
South Korea	Health Exp.	7.17	7.17	0.0
	GPI	0.549	0.549	0.0
	Physicians	2.14	0.73	−0.7
	Tourism index	0.20	0.20	0.0
Chile	Health Exp.	5.58	5.58	0.0
	GPI	0.629	0.445	−0.3
	Physicians	1.02	0.6	−0.5
	Tourism index	0.23	0.23	0.0
Mexico	Health Exp.	6.24	4.05	−0.4
	GPI	0.411	0.411	0.0
	Physicians	2.10	0.38	−0.8
	Tourism index	0.22	0.22	0.0
Costa Rica	Health Exp.	8.10	4.12	−0.5
	GPI	0.570	0.418	−0.3
	Physicians	1.11	0.39	−0.6
	Tourism index	0.23	0.23	0.0
Brazil	Health Exp.	3.80	3.80	0.0
	GPI	0.488	0.385	−0.2
	Physicians	1.9	0.4	−0.8
	Tourism index	0.23	0.23	0.0
India	Health Exp.	5.88	3.82	−0.4
	GPI	0.389	0.389	0.0
	Physicians	0.70	0.36	−0.5
	Tourism Index	0.24	0.24	0.0
Singapore	Health Exp.	3.20	3.20	0.0
	GPI	0.7	0.3	−0.5
	Physicians	1.95	0.30	−0.8
	Tourism index	0.19	0.19	0.0
Malesia	Health Exp.	7.28	4.36	−0.4
	GPI	0.635	0.442	−0.3
	Physicians	1.2	0.4	−0.7
	Tourism index	0.21	0.21	0.0
Philippines	Health Exp.	4.52	3.65	−0.2
	GPI	0.421	0.370	−0.1
	Physicians	1.15	0.35	−0.7
	Tourism index	0.25	0.25	0.0
United Arab Emirates	Health Exp.	3.20	3.20	0.0
	GPI	0.596	0.324	−0.5
	Physicians	2.53	0.30	−0.9
	Tourism index	0.21	0.21	0.0

19.5 Hypothesis Testing

In addition to DEA, hypothesis testing is used for inferential purposes. There, the aim is to reveal the relationship between the below mentioned indices with DEA scores. The main purpose of this section is to evaluate other input options that can be included in future DEA models for medical tourism performance.

Benefiting from the Global Innovation Index Report (World Economic Forum 2013), research hypotheses were constructed to determine the impacts of government effectiveness, regulatory quality, knowledge and technology outputs of countries on medical tourism efficiency. A brief description of tested indices is given below with short definitions:

Government effectiveness index *"An index that captures perceptions of the quality of public and civil services and the degree of their independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies".*

Regulatory quality index *"An index that captures perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private-sector development".*

Scientific and technical journal articles index *"The number of scientific and engineering articles. Article counts are from a set of journals covered by the Science Citation Index (SCI) and the Social Sciences Citation Index (SSCI)".*

Index values that belong to the top medical tourism destinations are presented in Table 19.4. The low and high index values for each hypothesis are determined according to the average of index values for each group of countries. The value that is above the average is taken into account as "high" index group. Independent-samples t-test is applied using SPSS-23 in order to test whether there is a statistically significant difference between groups for the significance level; $\alpha = 0.05$. The related hypotheses and test results are given below in Table 19.5. Each research hypothesis is tested against the null hypothesis that there is no difference between countries.

Hypothesis test results reveal that although direct interventions set by governments regarding medical tourism has a positive effect on efficiency, overall government managerial performance has no effect on medical tourism efficiency. Similarly, we find out that regulatory quality and the number of scientific and technical publications that belongs to the considered countries have no significant effect on medical tourism efficiency scores. It can be inferred from the results that regulatory quality, political and scientific effectiveness of countries have no direct effect on their medical tourism performances.

Table 19.4 Index scores of top medical tourism destinations

DMU	Governmental effectiveness	Scientific and technical publications	Regulatory quality
Argentina	**71.0**	10.3	21.5
Brazil	**65.3**	14.6	49.6
Chile	**79.8**	**17.4**	**87.2**
Costa Rica	**80.0**	10.3	30.4
Cuba	56.1	10.3	58.7
Hungary	57.7	**29.5**	**76.9**
India	36.7	9.6	35.2
Israel	34.3	**46.6**	**78.6**
Jordan	55.9	**26.5**	50.9
Malaysia	**69.4**	**15.9**	**64.2**
Mexico	49.1	5.9	60.0
New Zealand	**91.3**	6.0	**99.5**
Panama	**63.3**	5.9	57.7
Philippines	**68.0**	2.1	**93.0**
Poland	56.9	**25.7**	**74.7**
Singapore	**95.8**	**31.0**	**100**
South Africa	**66.6**	15.0	58.7
South Korea	**71.8**	**29.1**	**73.9**
Taiwan	**76.4**	**46.6**	**78.6**
Thailand	41.1	9.0	53.5
Turkey	43.4	**20.9**	59.1
U.Arab Emirates	**89.6**	4.3	**68.5**
Vietnam	30.8	5.4	33.6

19.6 Conclusions

The medical tourism industry has been growing worldwide. Due to the limited studies in the published literature, developing a methodology for evaluating the performance of medical tourism has become crucial. This study proposes a categorical DEA model in order to obtain a better decision model for evaluating medical tourism efficiency of countries.

We believe that the contribution of this study to the related literature is twofold. Despite the strong growth of DEA studies in the literature, there is no study concerning medical tourism. Additionally, this chapter also contributes to the current research on medical tourism by introducing a DEA model although medical tourism efficiency has been overlooked in earlier studies.

The applied methodology allows us to compare the performances that belong to top medical tourism destinations. We can also identify the countries that operate below their performances and related improvement suggestions.

Table 19.5 Results of statistical analysis

Research hypotheses	*t*-test		Statistical decision
	t	*p*	
H1: There is a difference between countries with low government effectiveness and high government effectiveness regarding medical tourism efficiency	0.404	0.693	Reject
H2: There is a difference between countries with low regulatory quality and high regulatory quality regarding medical tourism efficiency	0.295	0.773	Reject
H3: There is a difference between countries with low scientific and technical publications and high scientific and technical publications regarding medical tourism efficiency	1.233	0.241	Reject

The results demonstrated how DEA approach seems to be very promising in the field of medical tourism. The constructed model includes the determination of the impacts of political environment, regulatory environment and knowledge and technology outputs on medical tourism. Policy makers and managers of medical tourism can use the outcomes of the proposed framework to take strategic action involving resource planning, allocation, and utilization.

As for future work, improvements must be made in this field focusing on additional dimensions such as the relative price of medical operations, the qualifications of hospitals regarding their capacity and qualifications (JCI accreditation, other quality standards, advertisement, expenses, etc).

References

Alsharif MJ, Labonte R, Lu Z (2010) Patients beyond borders: a study of medical tourists in four countries. Global Soc Policy 10:315–335

Banker RD, Morey RC (1986) The use of categorical variables in data envelopment analysis. Manag Sci 32:1613–1627

Beladi H, Chao CC, Ee MS, Hollas D (2015) Medical tourism and health worker migration in developing countries. Econ Model 46:391–396

Bies W, Zacharia L (2007) Medical tourism: outsourcing surgery. Math Comput Model 46:1144–1159

Charnes A, Cooper WW, Rhodes E (1978) Measuring the Efficiency of Decision Making Units. EJOR 2:429–444

Chuang TC, Liu JS, Lu LYY, Lee Y (2014) The main paths of medical tourism: From transplantation to beautification. Tour Manag 45:49–58
Cook WD, Seiford LM (2009) Data Envelopment Analysis (DEA)- Thirty years on. Eur J Oper Res 192:1–17
Crooks VA, Turner L, Snyder J, Johnston R, Kingsburt P (2011) Promoting medical tourism to India: messages, images, and the marketing of international patient travel. Soc Sci Med 72:726–732
Dash U, Vaishnavi SD, Muraleedharan VR (2010) Technical efficiency and scale efficiency of district hospitals: a case study. J Health Manag 12:231–248
De Nicola A, Gitto S, Mancuso P, Valdmanis V (2014) Healthcare reform in Italy: an analysis of efficiency based on nonparametric methods. Int J Health Plann Manag 29:48–63
Fetscherin M, Stephano RM (2016) The medical tourism index: Scale development and validation. Tour Manag 52:539–556
Gautam S, Hicks L, Johnson T, Mishra B (2013) Measuring the performance of critical access hospitals in Missouri using data envelopment analysis. J Rural Health 29:150–158
Glinos IA, Baeten R, Helble M, Maarse H (2010) A typology of cross-border patient mobility. Health Place 16:1145–1155
Gok MS, Sezen B (2013) Analyzing the ambiguous relationship between efficiency, quality and patient satisfaction in healthcare services: The case of public hospitals in Turkey. Health Policy 111:290–300
Grail Research (2009) The rise of medical tourism. www.grailresearch.com. Accessed 10 Sep 2016
Hajialiafzali H, Moss JR, Mahmood MA (2007) Efficiency measurement for hospitals owned by the Iranian social security organization. J Med Syst 31:166–172
Hoof VW, Pennings G, Sutter P (2015) Cross-border reproductive care for law evasion: a qualitative study into the experiences and moral perspectives of French women who go to Belgium for treatment with donor sperm. Soc Sci Med 124:391–397
Hu HH, Qi Q, Yang CH (2012) Analysis of hospital technical efficiency in China: effect of health ınsurance reform. China Econ Rev 23:865–877
Institute for Economics and Peace (2015) http://economicsandpeace.org/wp-content/uploads/2015/06/Global-Peace-Index-Report-2015_0.pdf. Accessed 25 Sep 2016
Ketabi S (2011) Efficiency measurement of cardiac care units of Isfahan hospitals in Iran. J Med Syst 35:143–150
Lee CG (2010) Health care and tourism: evidence from Singapore. Tour Manag 31:486–488
Lunt N, Horsfall D, Hanefeld J (2016) Medical tourism: a snapshot of evidence on treatment abroad. Maturitas 88:37–44
Lunt N, Smith R, Exworthy M, Green S, Horsfall D, Mannion R (2011) Medical tourism: treatments, markets and health system implications: a scoping review. OECD, Paris, p 55
Mainil T (2012) Transnational Health Care and Medical Tourism: Understanding 21st-century patient mobility: towards a rationale of transnational health region development. ISBN: 978–90–5472-219-9 NRIT Media, Nieuwegein
Mark BA, Jones CB, Lindley L, Ozcan YA (2009) An examination of technical efficiency, quality, and patient safety in acute care nursing units. Policy Polit Nurs Pract 10:80–186
Medical Tourism Association (2013) M. T.A Medical tourism climate survey. www.medicaltourismassociation.com. Accessed 25 Sep 2016
Medical Tourism Association (2016) 2015-2016 Annual report. http://www.medical tourismassociation.com/userfiles/files/GHR_M_0037_ANNUAL_REPORT.pdf. Accessed 01 Sep 2016
Medretreat (2016) http://www.medretreat.com/destinations/thailand.html. Accessed 20 Sep 2016
Mohamad WN, Omar A, Haroni MS (2012) The moderating effect of medical travel facilitators in medical tourism. Procedia- Soc Behav Sci 65:358–363
Nayar P, Ozcan YA (2008) Data envelopment analysis comparison of hospital efficiency and quality. J Med Syst 32:193–199
Pan T, Chen WC (2014) Chinese medical tourists – Their perceptions of Taiwan. Tour Manag 44:108–112

Pocock NS, Phua KH (2011) Medical tourism and policy implications for health systems: a conceptual framework from a comparative study of Thailand, Singapore and Malaysia. Glob Health 7:12

Prakash V, Annapoorni D (2015) Performance evaluation of public hospitals in Tamil Nadu: DEA Approach. J Health Manag 17:417–424

Rozée VE, De La Rochebrochard E (2013) Cross border reproductive care human reproduction. Hum Reprod 28:3103–3110

Smith RM, Álvarez M, Chanda R (2011) Medical tourism: A review of the literature and analysis of a role for bi-lateral trade. Health Policy 103:276–282

Taskopru V (2014) Comparative analysis of classical data envelopment analysis and categorical data envelopment analysis models for energy efficiency, Master's thesis, Mimar Sinan Fine Arts University, Istanbul

Veerasoontorn R, Beise-Zee R (2010) International hospital outshopping: a staged model of push and pull factors. Int J Pharma. Healthc Mark 4:247–264

Viladrich A, Faust RB (2014) Medical tourism in tango paradise: The internet branding of cosmetic surgery in Argentina. Ann Tour Res 45:116–131

Wilson GW, Jadlow JM (1982) Competition, profitincentives, andtechnicalefficiency in theprovision of nuclearmedicineservices. Bell J Econ 13:472–482

World Economic Forum (2013) The global information technology report. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.362.6163&rep=rep1&type=pdf. Accessed 01 Sep 2016

Yu JY, Ko TG (2012) A cross-cultural study of perceptions of medical tourism among Chinese, Japanese and Korean tourists in Korea. Tour Manag 33:80–88

Zere E, Mbeeli T, Shangula K, Mandlhate C, Mutirua K, Tjivambi B, Kapenambili W (2006) Technical efficiency of district hospitals: evidence from Namibia using data envelopment analysis. Cost Eff Resour Allocation 4:5–14

Zhou Z, Zhao L, Lui S, Ma C (2012) A generalized fuzzy DEA/AR performance assessment model. Math Comput Model 55:2117–2128

Part VIII
Health Care Management

Chapter 20
Healthcare Human Resource Planning

John Pastor Ansah, Victoria Koh, Steffen Bayer, Paul Harper, and David Matchar

20.1 Importance of Healthcare Human Resource Planning

Healthcare accounts for a large share of public expenditure in many countries (Bloor et al. 2003); 60–70% of these expenditures are devoted to the human resources. Aimed at optimizing the size and makeup of human resources relative to care needs, healthcare human resource planning impacts population health, healthcare costs, operations and access to healthcare (Scott 2011). If the supply of health professionals is inadequate to meet the needs of a particular population, morbidity and mortality can increase (Anyangwe and Mtonga 2007; Starfield et al. 2005; Kirch et al. 2012). An oversupply of healthcare workers, on the other hand, would waste valuable resources (Hurst and Kelley Patterson 2014).

Recruitment policies have to be carefully tailored to meet complex future demands (OECD 2013). Societal trends such as population aging affect both the demand and supply of health workers. An aging population generally increases the demand for healthcare services due to an increased prevalence of chronic diseases

J.P. Ansah • V. Koh (✉) • D. Matchar
Health Services and Systems Research, Duke-NUS Medical School, 8 College Road, 169857 Singapore, Singapore
e-mail: john.ansah@duke-nus.edu.sg; victoria.koh@duke-nus.edu.sg; david.matchar@duke-nus.edu.sg

S. Bayer
Southampton Business School, University of Southampton, Highfield Campus, SO17 1BJ, Southampton, UK
e-mail: s.c.bayer@soton.ac.uk

P. Harper
School of Mathematics, Cardiff University, Senghennydd Road, CF24 4AG, Cardiff, UK
e-mail: harper@cardiff.ac.uk

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_20

Table 20.1 Taxonomy of Healthcare human resource planning problems

Problems	Description	Examples
Healthcare demand	Understanding the demand for healthcare services in order to provide adequate supply to meet demand	(Chung et al. 2010)
Healthcare human resource supply	Accounting for human resources required to meet healthcare needs	(Tomblin Murphy et al. 2014)
Attrition	Exploring the impact of different policies on the reduction of attrition and increase in work attractiveness	(Teye et al. 2015; Buchan and Sochalski 2004)
Skill mix	Finding an appropriate skill mix to increase productivity of the workforce and to meet healthcare needs	(Fulton et al. 2011; Buchan and Calman 2005)
Training requirements	Understanding the amount and type of skill set required to meet care needs	(Willis et al. 2013)
New models of care	Exploring different means to organize care provision to effectively utilize available human resources in order to meet care needs	(Anson 2000)

(Evashwick et al. 1984; Park 2014) while at the same time the healthcare system might also have to adapt to an aging workforce.

Moreover, the costs of training the health workers, especially physicians, is very high and is often heavily subsidized by public expenditure. Training an individual physician in the United Kingdom, for instance, is largely funded by the government, and can cost up to £500,000 (Centre for Workforce Intelligence 2014b). Furthermore, healthcare professionals have to undergo training for many years before obtaining licensure to practice. Due to this time lag between education and practice, adjustments made to the training intake of healthcare professionals at any point in time are only effective after a long time has lapsed. Hence, human resource planners need to pre-empt any changes in human resource requirements well in advance.

Planning generates insights into various healthcare human resource problems and facilitates the identification of leverage points for action. Table 20.1 provides a taxonomy of some healthcare human resource planning problems with references to selected literature.

20.2 Elements of Healthcare Human Resource Projections

Healthcare human resource planning requires taking the context of a particular healthcare system into account as well as the careful analysis of both supply and demand/need factors. The analysis of both past and current behavioral trends is useful as a basis for the projection of future requirements. This section describes the elements necessary in projections of human resources for health.

The projection of the supply of health professionals is essentially the sum of the current workforce and new hires less attrition. Four main factors act to increase or decrease the human resources available: specialized training requirements, worker preferences for type and site of service, attrition and migration, and service models.

Specialized Training Requirements Due to the highly specialized nature of healthcare human resources, training is the primary means to increase or decrease the supply of healthcare professionals. Thus, careful monitoring of the number of admissions to and graduations from training institutions such as medical and nursing schools is required. Application trends to specific healthcare professions (physicians, nurses and allied healthcare professions) and specialties must also be identified in order to understand if the supply of health professionals can adequately meet population needs. It is also important to distinguish between domestic and international students as the latter's decision to stay in the country of study to work upon graduation or return to their home country will need to be taken into account in admissions and recruitment policies (Hawthorne and Hamilton 2010).

Worker Preferences for Type and Site of Service Upon completion of education, graduates enter the health workforce as new hires in the various healthcare institutions across the system. Employment into the private or public sector and in the city or rural areas should be considered separately as factors such as age and gender have been found to influence workplace preference and can result in unequal distribution of healthcare human resources (Frank et al. 1997; Fu et al. 2013; Thompson et al. 2008; Kimball et al. 2004). An understanding of the geographical distribution of healthcare human resources is also important in managing imbalances and ensuring that there is adequate access to health care services everywhere. To address the issue of unequal healthcare human resource distribution, countries such as Thailand implement rural field residencies or internships as a requirement in medical training (Dussault and Franceschini 2006).

The advent of modern transportation has greatly facilitated the movement of healthcare workers across borders. Often, health professionals from low- and middle-income countries move to high-income countries in North America and Western Europe to seek employment (Aluttis et al. 2014). Some countries (for example Australia) actively recruit foreign health professionals in response to healthcare human resource shortages. Moreover, migration of healthcare human resources also includes labor movement from rural regions to city areas. These trends can exacerbate existing imbalances in the supply of health professionals in the developed and developing countries or rural and urban regions.

Attrition and Migration Various factors account for healthcare human resource attrition. Since the process of training human resources for health can be demanding, dropout rates from the programs are relatively high. Dropout rates have to be closely monitored and taken into account when planning admissions to training institutions. Health professionals can exit the labor force due to retirement, death, or other reasons. Naturally, retirement behavior of healthcare human resources is closely linked to its age distribution. Besides resignation from healthcare human resources

themselves, movement between different types of providers (e.g. public and private sector providers) or from rural regions to urban areas should also be taken into consideration when projecting future healthcare human resource requirements. Similarly, dismissals reduce the number of health professionals in the labor force. However, the same individual may seek employment in another institution within the healthcare system thus re-entering the labor force. Emigration and the movement between different regions of a country also have to be considered to gain a thorough understanding of the distribution of human resources.

It is worth noting that the availability of health professionals to provide clinical services is also a function of their time allocation: besides clinical work, health professionals may devote their time to administration, research or might be temporally on leave without permanently leaving the healthcare labor market. Therefore, a more accurate representation of the supply of health professionals is the full time equivalent of each profession available for clinical work, taking into account the workload of each individual.

Service Models The manner in which healthcare is provided, is dependent upon the organization of care delivery within a population, thereby affecting human resource requirements. Three main issues of organization of care delivery—task shifting, changes in service models and productivity—are discussed below.

First, task shifting is the delegation of tasks to a less specialized health professional wherever appropriate, to ensure the efficient use of available (human) resources (World Health Organization 2007). Enhancing the roles of health professionals that are less trained allows more highly trained staff to focus on tasks that are more complex. This strategy is used in many developing countries to resolve healthcare human resource shortages and to improve cost efficiency (Fulton et al. 2011). Healthcare human resource shortages may also be alleviated by changing the mix of skills within an organization to meet previously unmet care needs and improve the quality of care. Next, changes in service models also influence the manner in which healthcare services are provided and thus the demand for specific healthcare services. This can impact the requirements for healthcare human resources. Possible changes also include changes to the setting (such as a GP's office or a hospital) where specific services can be provided (e.g. such as enhancing primary care (Dubois and Singh 2009)). Lastly, productivity may be broadly defined as the quantity of health services provided per health professional and affects human resource requirements. It is important to note that longer work hours do not necessarily imply a higher workforce productivity. Another factor that can influence productivity is technological advancements.

20.3 Challenges of Healthcare Human Resource Projections

Uncertainty Although the elements of healthcare human resource planning are relatively straightforward, the actual task of projecting healthcare human resource requirements in the future is a complex process. To simplify the task, in general,

most forecasting models rely upon previously observed relationships among parameters and assume that these trends will persist into the future. However, there exists a large degree of uncertainty in regards to this assumption. The magnitude of this uncertainty will increase as the projection timeframe increases.

Availability of Detailed Data Another limitation of this historical approach is often the level of detail and disaggregation of the available data. Much of the data on demand for healthcare services are gathered through surveys and may be subject to reporting bias. Furthermore, some surveys are voluntary and response rates may be low. Also, data such as the age distribution of the healthcare human resources may not be well documented, especially for those in allied health professions. Some data might also not be available for planning purposes since providers might keep it confidential for commercial reasons.

Dynamic Complexity Healthcare systems are dynamic. Besides the constantly emerging new knowledge and techniques to improve the scientific aspect of healthcare, stakeholders are also continually looking to increase efficiency of the system. These advancements can change the way service is provided and also health seeking behaviors.

Task-Shifting Although task shifting, such as from doctors to nurses or nurses to community health workers, has proven to be effective in alleviating healthcare human resource shortages and increasing the efficiency of a healthcare system, the appropriate and acceptable extent of task shifting is often contested. Within a healthcare system, there are often widely different views about which services and activities can be allocated to less trained staff while assuring the quality of healthcare services. The extent to which task-shifting can actually be implemented is another source of uncertainty in healthcare human resource planning. It is difficult to determine the optimal level of skill mix in a healthcare system as it is highly dependent on a variety of parameters such as the amount of resources available, population needs and their preferences for service providers, cost, quality of care, and regulatory environment (Buchan and Calman 2005). The scope of practice of an individual health professional is limited by rules and regulation and might even differ between jurisdictions within a country (Green et al. 2013). The regulatory context and possible or likely changes to the allowed scope of practice of different healthcare professionals therefore have to be taken into account.

Lack of Standard Metrics for Productivity Quantifying worker productivity in healthcare is crucial in planning, but lack of standard metrics make this quantification difficult. Processes of care (e.g., average patient visit per diagnosis (Rhoads et al. 2006)) or workload (e.g., average patient length-of-stay (North and Hughes 2012)) are commonly used as measurements of healthcare human resource productivity (Maynard 2006). Technological advancements frequently increase productivity and reduce cost-per-unit of service (Girifalco 1991). This may be achieved through the reduction of manual labor required such as the automated electrocardiogram (EKG) readings for cardiologists, reducing the read time per

EKG (Lomas et al. 1985). New technologies often require specialized equipment and special training which must be accounted for in healthcare human resource planning.

20.4 Healthcare Human Resource Planning Frameworks

Four types of frameworks are commonly used in healthcare human resource planning (Goodman 1996; Scheffler et al. 2008; United States Bureau of Health Resources Development 1974): (a) needs-based; (b) utilization-based; (c) workforce-to-population ratio; and (d) service-target based. Each framework is conceptually unique with its own distinct set of assumptions and data requirements. Healthcare human resource projections obtained by different approaches can therefore differ significantly. Besides these typical frameworks, in practice, combined approaches have been developed (Dreesch et al. 2005; Centre for Workforce Intelligence 2014a). The decision to adopt any one of these frameworks is highly dependent on the context of each specific healthcare situation.

Needs-Based Framework The needs-based framework projects future healthcare human resource requirements based on the current estimated healthcare needs of a population. Healthcare needs herein refer to the healthcare services necessary to provide adequate healthcare to maintain a healthy population (i.e., to alleviate avoidable deterioration) and to care for individuals in that population who have health conditions (i.e., to cure or otherwise optimize health status). Elements considered in this framework are demographic characteristics such as disease prevalence, age, gender and education level of a population (Roberfroid et al. 2009). The needs-based framework is based on the assumption that all healthcare needs will be met; that economical methods to address these needs can be identified and implemented; and that healthcare resources will be employed relative to needs (Dreesch et al. 2005). The needs-based framework assumes that the demand for healthcare services is not price sensitive (Goodman and Weyant 1990). Often used as an advocacy tool, the projection of healthcare human resource requirements is seen by some as unobtainable and "idealized" as it assumes that the sole determinant of healthcare demand is medical need. Other variables such as socioeconomic status, healthcare financing, and increased screening policies have been shown to affect health-seeking behavior (Andersen and Newman 1973; Miller 2012; Van Der Heyden et al. 2003). The needs-based framework requires extensive epidemiological data, which is often unavailable.

Utilization-Based Framework Utilization-based, or demand-based, frameworks estimate the future healthcare human resource requirements using the current levels of services utilized by the population as a proxy for satisfied demand. Satisfied demand is defined as the levels of healthcare services a population will seek and have the ability to acquire at the current pricing within a certain timeframe.

Similar to the needs-based framework, demographic information such as disease prevalence, age, gender and education level are fundamental to this framework. In addition, utilization patterns of healthcare services and market factors that influence these patterns are also taken into consideration (Roberfroid et al. 2009). This framework relies on the following assumptions: current demand for healthcare is adequately satisfied by current level, combination and distribution of health services; age- and gender-specific requirements are held constant into the future; and demographic changes over time can be predicted based on prevailing trends (Birch et al. 2007). Due to the assumption that there is little or no change in the population-specific utilization patterns, the utilization based approach is useful in predicting economically feasible targets. It is often used in studies of geographical variations. This approach does not account for changes in future utilization patterns. In addition, information on the utilization and demand for healthcare services, especially in the private sector, is not always available. It must also be noted that the disparity between demand, utilization and needs for services is not taken into consideration in this approach (Dussault et al. 2010).

Workforce-to-Population Ratio Framework The workforce-to-population ratio framework requires the simple calculation of healthcare personnel per unit of population. This ratio is then compared to benchmarks or expert opinions to determine human resource adequacy. Elements considered in this approach are typically demographic data such as population growth, and workforce strength. To increase the sophistication of resultant projections, various adjustments have been made to account for factors such as utilization rates by age or gender and attrition rates of healthcare human resources (Cromwell et al. 1991; Weiner 1994). The workforce-to-population framework often uses a ratio from a reference country or region with a slightly more developed healthcare sector as a benchmark (Dreesch et al. 2005). Although the speed and ease of application is an apparent advantage of this framework, it does not take into account productivity, utilization and distribution of healthcare personnel.

Service-Target Based Framework In service-target based frameworks, health authorities or policy makers establish goals for the production and delivery of health services based on specific outcomes of interest. Taking into account staffing and productivity standards, these targets are then translated into healthcare human resource requirements (Hall and Mejia 1978). The elements crucial to this framework are information on the population and healthcare human resources such as demand for healthcare services and level of service provision respectively. The underlying assumption in the service-target framework is that the standards of each service covered are achievable within the projection timeframe (Dreesch et al. 2005). The service-target based framework is relatively easy to apply. Since services provided by healthcare human resources can be disaggregated, they can be matched with each component part of the health system with relative ease to estimate suitable targets. Moreover, in planning for health services provided by a team of health professionals, the skill mix of the system is accounted for. While

this framework considers productivity growth, there may be judgment errors about the ability of the healthcare sector to improve productivity during the projection timeframe, resulting in unrealistic targets.

20.5 Survey of Analytical Methods

This section explores some of the common analytical methods applied to healthcare human resource planning. They are: (a) system dynamics; (b) linear programming; (c) data-driven forecasting; (d) microsimulation and (e) econometrics.

20.5.1 System Dynamics

System dynamics is increasingly being applied to the development of healthcare human resource models to support planning of healthcare human resources. It is a method for depicting and simulating dynamic behavior of health, economic and social mechanisms, based on information feedback, and the concept of stock and flow, delays, and nonlinearity. The foundation of a system dynamics model is coupled, nonlinear, first order differential (or integral) equations. For brevity, the modeling process begins with defining a dynamic problem (something that changes over time), proceeds through mapping and modeling stages, to verification and validation of the model and its policy implications.

In forecasting human resource requirements, there has been a recognition of the intrinsic uncertainties and complexities of the factors influencing the demand and supply of health professionals. An apparent advantage of system dynamics is its ability to (a) engage stakeholders in model development to improve understanding and ownership, (b) represent different parts of the health system in one single framework to facilitate comprehensive analysis of policy impacts, (c) address nonlinear relationships and different scenarios and interventions simultaneously in a single framework. Nevertheless, compared to microsimulation and other methods, system dynamics applies top-down approach in modeling aggregate decision change process, using aggregate stocks to represent groups. System dynamics models can be very complex as the number of variables and casual relations increases. The list of data input can increase rapidly making it difficult to populate.

20.5.2 Linear Programming

Linear programming is an optimization technique used to solve decision problems involving a single objective function and its constraints which are linear functions of decision variables (Wang 2005). The theoretical goal of linear programming

is finding the set of positive decision variables that maximizes or minimizes the objective function. Linear programming problems can be solved by computer methods such as the *simplex* algorithm. The main advantage of this method is the simplification of a real world problem into one that can be described with a linear relationship. The resultant simplicity allows for model variations to explore what-if analyses. Typical applications of the linear programming method are the planning and allocation of healthcare human resources for a single worker class such as registered nurses (Lavieri and Puterman 2009). The strengths of this method is also its intrinsic weakness. By definition, linear programming can only solve decision problems with only one objective function. However, in practice, projecting requirements for healthcare human resources involves more than one outcome measure. Furthermore, specification of the output has to be clearly defined, which may not always be possible especially in such a dynamic system like healthcare.

20.5.3 Data Driven Forecasting

Data-driven forecasting involves the formulation of a series of equations to project human resource requirements. This method uses available data, such as patient visits, number of available physicians, work hours of physicians, and service uptake rate to predict the size of healthcare personnel required (Spetz et al. 2015; Tjoa et al. 2010; Hooker et al. 2011). This method tends to be static and often do not account for expected future changes. However, the major advantage of this method is its ease of application.

20.5.4 Microsimulation

The increasing availability of healthcare data and faster computers has led to recent development of microsimulation models for healthcare human resource planning. What distinguishes microsimulation from other methods is the bottom-up approach. For healthcare human resource planning, demand for services are determined at the individual level and vary substantially by individuals in a representative sample of the population; while on the supply side, it is posited that career decisions are taken by individuals and household, hence the need to represent physicians' career decisions individually. A unique attribute of microsimulation is the ability to account for individual's demographics, health-related behaviors, socioeconomic factors, and other risk factors; thus increasing the accuracy of estimation of demand for care. The strength of microsimulation is its ability to model paradigm shifts in care delivery and policies that have impact only on specific patient groups, while the major drawbacks are its complexity, greater data requirements and increasing processing time.

20.5.5 Econometrics

Econometrics is an amalgamation of statistics, mathematics and economics that takes into account the present and future constraints on resources to predict the healthcare human resources required for the healthcare sector. There are two types of econometric models: macro models and micro models. The former deals directly with aggregate variables while the latter deals with behavior of individuals or institutions. Thus, micro models are more suitable if distributive patterns are of particular interest. Although micro and macro models are conceptually different, the former can also be used to evaluate aggregate variables by application of the summation function. In an econometric model, demand and supply interact and converge at an equilibrium point. Factors such as pricing, met and unmet needs that influence the market supply and demand are integrated into this method to evaluate the demand for healthcare personnel (O'brien-Pallas et al. 2001; Yett et al. 1972). Econometric models are particularly useful for examining the relationships among stock, wages, demand, and budgets and are thus often used for evaluations on a state- or nation-wide scale (Feldstein and Roehrig 1980; Feldstein 1967; Yett et al. 1975). Since this method mainly focuses on market factors that influence labor participation and healthcare utilization, other factors such as population health needs, impact of government policy, the influence of the changing health system and the impact of outcomes are not adequately accounted for (O'brien-Pallas et al. 2001). Moreover, econometric models generally require extensive amounts of data which may often not be available.

A survey of healthcare human resource planning applications is shown in Table 20.2 below. The examples listed are chosen to represent the different contexts in which the various planning frameworks and analytical methods can be used.

Table 20.2 Survey of applications

Paper title and author	Approach and method	Aims/question/ objective	Assumptions/ framework	Summary of results
Needs-based approach				
Forecasting the need for medical specialists in Spain: application of a system dynamics model (Barber and López-Valcárcel 2010)	System dynamics	To plan for healthcare human resources, and to simulate changes in supply and demand of physicians in Spain	Specialist requirements based on Delphi-type consultation of experts Baseline model assumes no change to controllable parameters	The shortage of doctors in certain specialties and rural areas will worsen over the years

(continued)

Table 20.2 (continued)

Paper title and author	Approach and method	Aims/question/ objective	Assumptions/ framework	Summary of results
A novel approach for the accurate prediction of thoracic surgery workforce requirements in Canada (Edwards et al. 2014)	Micro-simulation	To develop a microsimulation model of thoracic surgery workforce to forecast future supply and demand	No interaction between supply and demand. The number of thoracic surgeons do not influence the incidence of lung cancers	As the incidence of lung cancer is projected to decrease after 2030, the current shortage of thoracic surgeons will reverse and become an excess supply in future
Utilization-based approach				
Forecasting the absolute and relative shortage of physicians in Japan using a system dynamics model approach (Ishikawa et al. 2013)	System dynamics	To estimate future physician-supply using system dynamics modeling	Constant rate of retirement, deaths and movement across specialty. No change to healthcare system	The number of physicians increases during 2008–2030 and the shortage would resolve at 2026 for all clinical physicians except for OB/GYN
Supply and demand analysis of the current and future US neurology workforce (Dall et al. 2013)	Micro-simulation	To forecast the supply and demand for neurologists through 2025 nationally and by state	Supply: Current retirement, work hours and recruitment unchanged. Demand: 10% and 20% shortfall of adult and child neurologists respectively	The shortfall of neurologists is likely to persist under all scenarios
Workforce-to-population ratio approach				
Theoretical system dynamics modeling for Taiwan pediatric workforce in an era of national health insurance and low birth rates (Wu et al. 2013)	System dynamics	To predict the gaps between needs and supply of pediatric workforce in the future and examine possible intervention attempts	Only the size of the pediatric population and the expected quality of pediatric care influence workforce needs. Infant mortality of 4.0/1000 live births	Under the base scenario, shortage of pediatricians fluctuates. Under the MDG scenario, there is a consistent huge shortage of pediatricians

(continued)

Table 20.2 (continued)

Paper title and author	Approach and method	Aims/question/ objective	Assumptions/ framework	Summary of results
Projection of the dental workforce from 2011 to 2020, based on the actual workload of 6762 dentists in 2010 in Taiwan (Huang et al. 2013)	Data-driven	To project the dental workforce from 2011 to 2020 in Taiwan using results from a survey of dentists	A dentist-to-population ratio of 5.0 is adequate	There will be a surplus of dentists in 2020 due to a dwindling population
Service-target approach				
Optimizing nursing human resource planning in British Columbia (Lavieri and Puterman 2009)	Linear programming	To model the Registered Nurses workforce in British Columbia	Service needs are known. Constant age distribution of nurses recruited and promoted. At least 500 nurses must be recruited annually	Current policy is not sustainable. Age distribution had a significant impact. In all cases, a higher recruitment level was required
Scaling up priority health interventions in Tanzania: the human resources challenge (Kurowski et al. 2007)	Data-driven	To explore the coverage of priority interventions to achieve Millennium Development Goal targets in mainland Tanzania. Healthcare human resource availability for 2015 were forecasted and compared	Constant disease incidence, prevalence and health risks. Alternative scenarios to investigate the impact of different policies on requirement of health professionals	The scaling up of the priority interventions would increase the requirements for health professionals, which are unlikely to be available in mainland Tanzania

20.6 Considerations for Selecting Healthcare Human Resource Planning Approach

First, the demographic characteristics of a population to be served and its possible changes over time should be identified. In general, a growing population is expected to demand more healthcare services. However, the increase in demand between

young and aging populations will differ. A young population is expected to be relatively healthy with disease burdens unlikely to change drastically, resulting only in a proportionate increase for healthcare services. In contrast, the prevalence of chronic conditions in an aging population is expected to increase, resulting in a more than proportionate increase in the demand for healthcare services (Evashwick et al. 1984). This has implications for the number of healthcare workers required to maintain a healthy population. In addition, demographic distribution such as gender and ethnicity must also be taken into consideration as disease prevalence has been shown to differ among these categories (Zhang et al. 1990; Loh et al. 2015; Park and Kim 2015). The workforce-to-population ratio simply projects a ratio that does not account for changes in demographic characteristics, thereby yielding a significantly underestimated requirement for healthcare human resources. Thus, when changes in demographic characteristics are expected, the other three approaches—needs-based, utilization-based, and service-target based—may be more appropriate because these approaches account for changes in demographic characteristics.

The duration over which the human resource requirement is projected is also an important consideration as the parameters considered in the projection of healthcare demand and supply are subject to changes over time. In short timeframes where changes to the factors are not expected, the relatively simply workforce-to-population ratio approach may be used.

Healthcare utilization is dependent on a variety of factors. The socioeconomic factors that influence healthcare utilization have been well documented (Agerholm et al. 2013; Van Der Heyden et al. 2003). Higher educational attainment is expected to positively influence healthcare utilization patterns. Studies have shown that those with better education have increased expectations of their health status and thus are likely to have a higher number of preventive visits (Hulka and Wheat 1985; Machry et al. 2013). An increasingly educated population is also likely to be more affluent suggesting an increased ability to afford health care services (Cooper et al. 2012) in systems with high out-of-pocket costs. If utilization is expected to change over time, the selected approach should account for changes in expected utilization for projecting future requirements for healthcare human resources.

Another important factor to be considered when selecting a suitable forecasting approach is the mode of healthcare financing within a population. Whether healthcare is supplied by free market mechanisms, public provision or paid for under insurance coverage affects utilization rates and in turn healthcare human resource requirements. Since socioeconomic status is one of the determinants of healthcare utilization, the utilization-based approach may be more suited for health workforce forecasting in a healthcare system primarily funded by out-of- pocket payment. On the other hand, the needs-based approaches may be more suitable for forecasting in a universal healthcare system or one with compulsory social insurance as it aids in the identification of the minimum number of healthcare personnel required to attend to the health needs of everyone in a population (Wendt 2009).

Lastly, the productivity of the healthcare workforce is influenced by a myriad of factors such as technology, care organization, and new models of care. Any increase or reduction in workforce productivity will significantly affect the projection

numbers and therefore cannot be neglected. The increasing use of sophisticated technology in healthcare has been shown to increase productivity, reducing human resource requirements. However, it must be noted that whilst the use of technology often serves to increase productivity, in some cases the opposite effect might be seen as machine operation may require constant supervision or manual input. In addition, the manner in which healthcare is organized can have an effect on health workforce productivity. In eye care, for example, an increasing number of activities (e.g. fundus photography and management of diabetic retinopathy patients or intraocular pressure management in glaucoma patients) could potentially be performed by non-doctors (e.g. optometrist and ophthalmic technicians). Thus there is a need to optimize the skill mix to deliver healthcare efficiently to a population.

Healthcare human resource forecasting under different approaches is likely to produce both significant differences and similarities. The appropriateness of the forecasting approach used depends on the changing characteristics of the population to be served, the timeframe of the forecast, how factors influencing utilization of care are expected to change over time, healthcare financing and how the productivity of the workforce is likely to change.

20.7 Case Study 1: Future Requirements for and Supply of Ophthalmologists for an Aging Population in Singapore (Ansah et al. 2015)

20.7.1 Background

This case study demonstrates the process of synthesizing data from multiple sources to develop a strategic eye care workforce planning model, based on the system dynamics methodology that links population eye care needs to ophthalmologists' requirements in order to generate evidence-based projections to inform policy. The aim of this research was to estimate the future requirements for and supply of ophthalmologists in Singapore, up to year 2040, considering a range of plausible scenarios. Here we present only the highlights of the findings; details have been presented elsewhere (Ansah et al. 2015).

An assessment of the eye care workforce in Singapore is timely due to a plethora of factors that can significantly affect both the future demand for eye care services and the supply of eye care workforce. Singapore's aging population is likely to see a rise in the number of individuals with chronic eye conditions as the prevalence of several of such conditions has been shown to increase with age (Wong et al. 2006). In Singapore, those 65 and above is expected to increase by twofold from 2010 to 2050 (Asher and Nandy 2008). Population aging, combined with population growth and increasing life expectancy, is expected to result in a substantial increase in eye diseases and demand for eye care. Apart from potential changes in eye care service demands due to an increasingly educated population with greater expectations and awareness of services provided, technological (e.g.

treatment techniques) and organizational innovations that serve to increase the efficiency of healthcare delivery can also influence demand. The epidemiological shift towards chronic eye conditions also implies a need to tailor the composition of the eye care workforce and their required skill sets.

Demand and workforce projections are also necessary in policy planning to address "unmet needs" for eye care in Singapore (Zheng et al. 2013). Furthermore, workforce training needs should be planned in advance as many healthcare professionals, such as ophthalmologists, require long training periods before entering the workforce.

20.7.2 Methods

The Singapore Eye Care Workforce Model was developed using system dynamics and is a continuous time compartment model with explicit workforce stocks (Forrester 1961; Homer and Hirsch 2006; Sterman 2000). The model consists of interacting sets of differential and algebraic equations formulated using empirical data obtained from multiple sources. As discussed in the survey of methods and applications section, systems dynamics as a method is adept in addressing the inherent dynamic complexity within health workforce planning (Ishikawa et al. 2013; Masnick and Mcdonnell 2010). For specific application of system dynamics methods to workforce planning see the following references as cited (Ishikawa et al. 2013; Barber and López-Valcárcel 2010; Senese et al. 2015).

In developing the model, first a meeting with key stakeholders was organized to define the problem to be modeled. The problem definition includes defining reference modes—behavior overtime graphs—of key model variables that should be included in the model structure. In addition, the dynamic hypotheses that drive the behavior of the key variables were elicited from stakeholders and derived from the secondary literature to support the posited relationships. After that, a conceptual stock and flow computer model that captures the dynamic hypotheses elicited from the stakeholders, and can simulate the current behavior pattern of key variables was developed. This conceptual model was then presented to the stakeholders, comprising ophthalmologists, nurses, healthcare planners and managers, and health educators, to verify the soundness of the model. Following verification, the model was simulated using empirical data. Estimates from experts were used in place of unavailable data. After initialization, model validation was conducted by comparing simulated results to available time-series data (Homer 1996). During the model validation process, model refinements such as structure adjustment, level of aggregation and detail formulation were implemented. The iterative process of refinement and testing continued until the model outputs, structure and assumptions were accepted by the stakeholders to be a reasonable representation of the eye care system in Singapore. Finally, projections were generated under the base-case and alternative scenarios and presented to the stakeholders and the strategic workforce planning team.

20.7.3 Singapore Eye Care Workforce Model

There are three linked modules in the model: the prevalence of eye disease module, the demand module and the ophthalmologist requirement and supply module. Details of the model have been presented elsewhere (Ansah et al. 2015).

20.7.4 Scenarios

Based on the concerns expressed by stakeholders, the following scenarios were developed:

Business-as-Usual (BAU) In this scenario, key variables are assumed to be unresponsive to policy changes, i.e., uptake factor of eye services, current model of care, subsidies and workload of eye care workforce remain constant at the 2013 values throughout the projection timeframe. Using calibration, the estimated uptake factor was 4.5% for individuals with no education, 7% for those with primary education, 7.6% for those with secondary education, and 15% for individuals with tertiary education. Furthermore, the BAU assumes that only 5% of all patients with diabetic retinopathy (DR), glaucoma, myopia and refractive error, participate in the new primary care clinic initiative (PEC) and are referred from specialist outpatient clinics (SOCs) to PECs to be cared for by non-specialists. This scenario is not plausible in reality as demographic and policy changes are likely to influence uptake factor. Instead, it will serve as a reference point to evaluate the alternative scenarios.

Current Policy Similar to BAU, key variables with the exception of uptake factor among individuals with eye condition not receiving eye care are assumed to remain constant in the current policy scenario. This reflects the imminent increase in service utilization due to increased awareness, frequency of eye screenings, rising number of eye clinics in the community, new models of care incorporating new technology which makes eye care more accessible and available. Thus, it is assumed that individuals with untreated eye condition who seek care increase from 4.5% to 13% by 2040 for those with no education, while that for individuals with primary, secondary and tertiary education are 20%, 21% and 46%, respectively.

New Model of Care This scenario resembles the current policy scenario but accounts for an increased rate of patients in the PEC initiative. Twenty percent of all patients with DR, and glaucoma, as well as 90% of patients seeking care with myopia and refractive error, are assumed to be decanted from SOCs to PECs for follow-up with non-specialists (Gray et al. 2000; O'connor et al. 2012). Currently, there are very few clinics run by non-specialists (e.g. optometrists) in Singapore, but experiences from other countries such as U.K. and Australia provide evidence for the potential towards larger transitions.

Moderated Workload The moderated workload scenario assumes a 15% reduction in ophthalmologists' clinical duties, due to efforts to pursue non-clinical goals (such as research and education), improve work-life balance, and improve patient care, in addition to the adoption of the PEC initiative as described in the new model of care scenario.

20.7.5 Key Findings

The prevalence of eye diseases among Singaporean residents 40 years old and over are projected to more than double by 2040 with DR, glaucoma, and epiretinal membrane estimated to increase the most. The combined effect of an increasing and aging Singapore population explain the projected increase in the number of Singaporeans with eye diseases as prevalence of eye conditions generally increases with age.

Aside from an expected increase in prevalence of eye diseases, other factors such as increasing healthcare accessibility, frequency of eye screenings, and rising educational attainment are likely to result in an increase in the demand for eye care services. In addition, the local government has made efforts to increase accessibility and affordability of healthcare for the elderly, through policies such as the Pioneer Generation Package that offers additional subsidies on healthcare and insurance (Government of Singapore 2014). Where eye care is concerned, the Singapore government is supporting endeavors such as the launch of a Mobile Eye Clinic, which aims to provide comprehensive eye care for senior citizens who are unable to visit eye clinics due to physical or logistical restrictions (Standard Chartered Bank 2014). Similarly, the Singapore National Eye Centre (SNEC) started operations in various satellite eye clinics to reduce both traveling and waiting times for patients (SINGHEALTH 2008). These measures are likely to increase utilization of eye care services. Moreover, the general educational attainment of the future elderly population is expected to be higher due to legislation and affluence. In 2002, 57.1% of males and 51.6% of females in Singapore had at least a secondary education. In 2012, the proportions of males and females with at least a secondary education rose to 70.8% and 64.9% respectively (Teo 2013). The changing educational composition is likely to lead to higher visual acuity expectations, increasing further the utilization of eye care services.

All the scenarios considered project an increase in public sector ophthalmologists' workforce requirements by 2040 due to the increased demand for eye care services. To accurately inform ophthalmology training needs, the number of new ophthalmologists required to meet the projected demand was estimated, accounting for attrition. Under the BAU scenario, the projected number of ophthalmologists required by the year 2040 will increase 117% from 141 (sensitivity analysis at 95% confidence range: 125–216) in 2015 to 305 (227–436) by the year 2040. The current policy scenario projects a requirement of 144 (131–178) ophthalmologists

in 2015 and 396 (351–414) in 2040, representing a 175% increase in the number of ophthalmologists required, or 1.30 times that of the BAU scenario projection in 2040. Under the new model of care scenario, the public sector is projected to require 144 (131–179) ophthalmologists in 2015 and 359 (312–378) in 2040, increasing by 150%, and is 1.18 times that of the BAU scenario in 2040. Under the moderated workload scenario, 145 (131–181) ophthalmologists are projected to be required in 2015 and 422 (367–445) in 2040, which is a 192% increase and 1.38 times as many as the BAU scenario in 2040.

20.7.6 Insights from the Case Study

In planning for future eye care workforce requirements, it is imperative to consider the reasons for and magnitude of change in the number of people with eye diseases, demand for eye care services as well as workforce requirements. This is because changes in demand have direct repercussions on the ability of the healthcare system to adequately and effectively provide eye care services for an aging and growing population. In this study, these were accounted for through the use of an integrated planning framework that combines the utilization based approach and patient wait list, and anticipating future changes in the population and their demand for eye care services. The system dynamics methodology was adopted to simulate the dynamic and complex interactions in the eye healthcare system and to actively engage stakeholders in order to explore potential developments and issues that may influence workforce requirements over time.

Nevertheless, this projection also has limitations. One example is the use of the assumption that the increase in public sector visits for all patients will be proportional to the increase in residents aged 40 and over with eye conditions. Major changes in the proportion of visits by foreigners or by younger individuals would invalidate this assumption; however, foreigners and younger people represent a small proportion of all public sector visits. In addition, the projected demand depends on the projected demographic change in Singapore. Any significant changes observed in the population trend can affect the simulation results.

System dynamics is a highly suitable methodology for human resource planning in a dynamic context, where changes in the population or healthcare system are expected. This should be combined with a framework that minimizes unreasonable assumptions, and can be done in conjunction with stakeholders who must ultimately take ownership of the results in order to formulate compelling policy recommendations.

20.8 Case 2: Using Linear Programming to Optimize the Dental Team Skill-Mix for England (Harper et al. 2013)

20.8.1 Background

As oral health needs become increasingly complex with age, it is of import to adequately plan the provision of dental care for older people. Population aging, together with trends toward improvements in tooth retention, has a significant impact on the need and demand for dental treatment. In order to meet the growing needs for dental care the roles of different members of the dental team are being reconsidered. Besides dentists, dental hygienists, dental nurses, dental therapists, hygienists/therapists and clinical dental technician (who perform examinations of edentate (toothless) patients and provide patients with complete dentures) provide dental care.

To investigate dental skill mix requirements in order to meet future needs and demands of older people in England, a spreadsheet model consisting of three sub models—demand for dental treatment, workforce supply and optimization of workforce skill-mix—was developed. Workforce shortages or surpluses were determined by comparing estimated demand (expressed in costs and time) for various dental treatments with predicted workforce supply (accounting for workforce competencies and skill-mix). An optimization approach was then used to inform the planning of the workforce size and skill-mix to adequately meet future demand.

20.8.2 Demand Model

Both dynamic parameters such as demographic and edentulous rates, and static model inputs such as attendance and treatment rates were used in the demand model. Population forecasts for older adults (65+) were taken from government projections. Trends in increased tooth retention (partly due to fluoride in toothpaste) were represented in the sub-model to account for the reduction in need for complete dentures and the increase in demand for complex treatment among older people. Monte Carlo simulation was used to estimate the uncertainty surrounding projected demand due to uncertainties presented by future edentate rates.

National Health Service (NHS) participation refers to the proportion of the population who received treatment under the NHS. Future trends in NHS participation were taken into account using Monte Carlo sensitivity analysis.

Total demand for each treatment category and overall demand at various years were estimated by applying participation, attendance and treatment rates to the projected edentate and demographic rates in the model.

20.8.3 Supply Model

Current and future workforce shortages or surpluses were obtained by applying the estimated demand for treatment (expressed in hours) to workforce projections taking into account assumptions regarding clinical workload, level of NHS commitment (percentage of time spent by staff in each category on NHS work) and workforce full-time equivalents (FTEs).

20.8.4 Optimization of Workforce Skill Mix

We used linear programming to optimize the workforce skill mix by using information on competencies and costs, with treatment hours per care category. The linear program was run three times (for 2005, 2008 and 2028) to optimize the skill mix for the current, short and long term horizons.

Demand, expressed in hours and disaggregated by treatment category, was obtained from the demand model outputs for 2005, 2008 and 2028. Dental professionals within the research team indicated current and future dental treatments that will be performed in light of the introduction of the new staff roles (dental therapists/hygienists and clinical dental technicians). A range of scenarios were formulated by the experts to explore future competencies.

The linear program formulation minimizes a cost function subject to satisfying the demand for treatment hours for each of the treatment categories.

Inputs

d_i Is the demand for treatment hours for treatment category i, $i = 1, \ldots, 18$.

m_{ij} Is the limit on the number of treatment hours contributed by staff type j, to category i, $j = 1, \ldots, 5$, i $= 1, \ldots, 18$.

c_j Is the hourly cost of staff type j, $j = 1, \ldots, 5$.

h_j Is the conversion factor (a fraction) to convert treatment hours to required staff numbers for staff type j, $j = 1, \ldots, 5$.

Decision Variable:

x_{ij} Is the number of treatment hours contributed by staff type j to treatment category i , $j = 1, \ldots, 5$, $i = 1, \ldots, 18$.

Objective function:

Minimize $\sum_{j=1}^{5} c_j h_j \sum_{i=1}^{18} x_{ij}$

Subject to.

$$\sum_{j=1}^{5} x_{ij} \geq d_i \qquad i = 1, \ldots, 18. \tag{20.1}$$

$$x_{ij} \leq m_{ij} \qquad j = 1, \ldots, 5, i = 1, \ldots, 18. \tag{20.2}$$

$$x_{ij} \geq 0 \qquad j = 1, \ldots, 5, i = 1, \ldots, 18. \tag{20.3}$$

Constraint (20.1) specifies that demand for treatment hours is met and constraint (20.2) ensures that staff do not contribute more treatment hours than they are capable of. Constraint (20.3) ensures non-negativity.

The optimization module has two key outputs; the total staff for each category required to meet the needs of the population and the total annual cost of the number of staff.

20.8.5 Results

20.8.5.1 Demand Model

Due to population growth, treatment demand and costs for older people (65–99 years) are expected to see an absolute increase in the future. The model predicts that demand will increase more than 80% while associated treatment costs will rise by 250%. Where uncertainties around parameters were considered, results from the Monte Carlo simulation suggest that predicted demand may deviate from the average by 22%. Sensitivity analysis suggests that demand for treatment is particularly sensitive to dental attendance.

20.8.5.2 Comparison of Supply and Demand

Projections of workforce shortage and surplus were determined by comparing results from both the demand and supply model. Based on the current average demand, nearly a 100% NHS FTE would be required to for the workforce supply to meet demands, which is unlikely to be the case. Even after accounting for the uncertainty surrounding the precise magnitude of the workforce shortfall, there is an apparent shortage of dental staff in England initially, which may decrease or even disappear in the future.

20.8.5.3 Skill-Mix Optimization Model

The linear program was designed to facilitate the testing of scenarios involving various combinations of skill-mix and other factors.

Baseline (Scenario 1) Staff competencies were assumed to be based on the treatments that were most likely to be performed in the profession. Dental ther-

apists/hygienists were expected to be able to take on higher treatment workload and clinical dental technicians would be able to perform complete denture work by 2008. Results suggest that the supply of dental therapists/hygienists will experience the largest shortage, at 50%, while dentists will experience an oversupply of 23% by 2028.

No Skill-Mix (Scenario 2) Only dentists provide all treatment in this scenario and was evaluated to provide a reference point for comparison of the cost-benefits of skill-mix. Unsurprisingly, this scenario generated the highest staff costs (−19% for 2028).

Examinations by Clinical Dental Technicians and Hygienist/Therapists (Scenario 3) This scenario applies similar competencies to the baseline scenario. In addition, dental hygienists/therapists are assumed to perform routine and extensive examinations while clinical dental technicians will conduct examinations in edentate patients from 2008. Results suggest an unrealistic reduction in the number of dentists required due to the shift in responsibilities.

Clinical Dental Technicians Expand Their Role to Include Partial Dentures and Repairs (Scenario 4) Here, clinical dental technicians take on an expanded role to perform all denture work instead of only complete dentures as outlined in the baseline scenario. Total workforce requirements are estimated to be less than all previous scenarios due in part to the assumption that clinical dental technicians will spend 90% of their time providing care for older people whereas the other staff types are estimated to spend a significantly less amount of their time treating older patients. In 2008, this scenario generates a 33% saving in staff costs compared with the baseline scenario, underscoring the importance of introducing clinical dental technicians to the dental workforce.

Maximum Skill-Mix (Scenario 5) This scenario assumes a combination of the staff competencies used in scenarios 1, 3 and 4 to maximize staff capabilities. When compared to scenario 2 where no skill-mix was considered, results from this scenario clearly indicate the cost benefits of skill-mix. Cost-savings of about 44% relative to baseline may be generated by 2028 if a dramatic shift in workforce skill-mix is possible.

20.8.6 Recommendations

In spite of the uncertainty surrounding the level of NHS commitment and the FTE dental workforce, there was still an apparent shortage in dental workforce supply in the short-term and underutilization of both hygiene/therapists and clinical dental technicians. From the baseline scenario, it is recommended that additional NHS dental staff including 130 clinical dental technicians and as many extra hygienists/therapists as possible up to a maximum of 3400 in NHS FTEs be recruited immediately to meet the average shortfall. Although it is highly unlikely

that such large numbers of hygienists/therapists can be recruited in the short-term, the results of the skill-mix optimization clearly demonstrate the cost benefits of employing substantially more hygienists/therapists and also the introduction of clinical dental technicians to the publicly funded dental workforce. This has implications for training and longer term workforce capacity, together with developing and extending their roles.

Optimization results suggest that the most cost effective option may be significantly increasing the supply of hygienists/therapists and clinical dental technicians serving older people in order to maximize the skill mix of the dental team.

20.9 Future Research Directions

Healthcare human resource planning is a dynamic and complex issue. Despite the multitude of approaches and methodologies available, there has yet to be consensus on a definite technique for use in healthcare human resource planning. Strategic planning will require the use of an integrated approach that considers factors influencing both demand and supply of healthcare professionals, together with an appropriate analytical method that can incorporate less tangible factors such as skill mix and productivity in a single overarching framework.

References

Agerholm J, Bruce D, Ponce De Leon A, Burstrom B (2013) Socioeconomic differences in healthcare utilization, with and without adjustment for need: an example from Stockholm, Sweden. Scand J Public Health 41:318–325

Aluttis C, Bishaw T, Frank MW (2014) The workforce for health in a globalized context – global shortages and international migration. Glob Health Action 7:23611

Andersen R, Newman JF (1973) Societal and individual determinants of medical care utilization in the United States. Milbank Memorial Fund Q Health Soc 51:95–124

Ansah J, De Korne D, Bayer S, Pan C, Jayabaskar T, Matchar D, Lew N, Phua A, Koh V, Lamoureux E, Quek D (2015) Future requirements for and supply of ophthalmologists for an aging population in Singapore. Hum Resour Health 13:86

Anson BR (2000) Taking charge of change in a volatile healthcare marketplace. Hum Resour Plan 23:21

Anyangwe SCE, Mtonga C (2007) Inequities in the global health workforce: the greatest impediment to health in sub-Saharan Africa. Int J Environ Res Public Health 4:93–100

Asher MG, Nandy A (2008) Singapore's policy responses to ageing, inequality and poverty: an assessment. Int Soc Secur Rev 61:41–60

Barber P, López-Valcárcel BG (2010) Forecasting the need for medical specialists in Spain: application of a system dynamics model. Hum Resour Health 8:24–24

Birch S, Kephart G, Tomblin-Murphy G, O'brien-Pallas L, Alder R, Mackenzie A (2007) Human resources planning and the production of health: a needs-based analytical framework. Can Public Policy 33:1–16

Bloor K, Maynard A Hall J (2003) Planning human resources in health care: towards an economic approach: an international comparative review. Canadian Health Services Research Foundation= Fondation canadienne de la recherche sur les Services de santé, CANADA
Buchan J, Calman L (2005) Skill-mix and policy change in the health workforce: nurses in advanced role. OECD health working papers no. 17, OECD, FRANCE
Buchan J, Sochalski J (2004) The migration of nurses: trends and policies. Bull World Health Organ 82:587–594
Centre for Workforce Intelligence (2014a) Horizon 2035: health and care workforce futures Progress update
Centre for Workforce Intelligence (2014b) Robust workforce planning framework: an introduction. CfWI technical paper series no. 0001
Chung SH, Jung DC, Yoon SN, Lee D (2010) A dynamic forecasting model for nursing manpower requirements in the medical service industry. Serv Bus 4:225–236
Cooper RA, Cooper MA, Mcginley EL, Fan X, Rosenthal JT (2012) Poverty, wealth, and health care utilization: a geographic assessment. J Urban Health 89:828–847
Cromwell J, Rosenbach ML, Pope GC, Butrica B, Pitcher JD (1991) CRNA manpower forecasts: 1990-2010. Med Care 29:628–644
Dall TM, Storm MV, Chakrabarti R, Drogan O, Keran CM, Donofrio PD, Henderson VW, Kaminski HJ, Stevens JC, Vidic TR (2013) Supply and demand analysis of the current and future US neurology workforce. Neurology 81:470–478
Dreesch N, Dolea C, Dal Poz MR, Goubarev A, Adams O, Aregawi M, Bergstrom K, Fogstad H, Sheratt D, Linkins J, Scherpbier R, Youssef-Fox M (2005) An approach to estimating human resource requirements to achieve the Millennium Development Goals. Health Policy Plan 20:267–276
Dubois CA, Singh D (2009) From staff-mix to skill-mix and beyond: towards a systemic approach to health workforce management. Hum Resour Health 7:87
Dussault G, Buchan J, Sermeus W, Padaiga Z (2010) Assessing future health workforce needs. World Health Organization, Copenhagen
Dussault G, Franceschini MC (2006) Not enough there, too many here: understanding geographical imbalances in the distribution of the health workforce. Hum Resour Health 4:12
Edwards JP, Datta I, Hunt JD, Stefan K, Ball CG, Dixon E, Grondin SC (2014) A novel approach for the accurate prediction of thoracic surgery workforce requirements in Canada. J Thorac Cardiovasc Surg 148:7–12
Evashwick C, Rowe G, Diehr P, Branch L (1984) Factors explaining the use of health care services by the elderly. Health Serv Res 19:357–382
Feldstein MS (1967) An aggregate planning model of the health care sector. Med Care 5:369–381
Feldstein PJ, Roehrig CS (1980) A national econometric forecasting model of the dental sector. Health Serv Res 15:415–432
Forrester JW (1961) Industrial dynamics. M.I.T. Press, Cambridge, MA
Frank E, Rothenberg R, Brown WV, Maibach H (1997) Basic demographic and professional characteristics of US women physicians. West J Med 166:179–184
Fu MC, Buerba RA, Gruskay J, Grauer JN (2013) Longitudinal urban-rural discrepancies in the US orthopaedic surgeon workforce. Clin Orthop Relat Res 471:3074–3081
Fulton BD, Scheffler RM, Sparkes SP, Auh EY, Vujicic M, Soucat A (2011) Health workforce skill mix and task shifting in low income countries: a review of recent evidence. Hum Resour Health 9:1–1
Girifalco LA (1991) Dynamics of technological change. Springer US, Boston
Goodman DC (1996) Benchmarking the US physician workforce. An alternative to needs-based or demand-based planning. JAMA 276:1811–1817
Goodman HS, Weyant RJ (1990) Dental health personnel planning: a review of the literature. J Public Health Dent 50:48–63
Government of Singapore (2014) Pioneer generation package – overview [Online]. Available: http://www.cpf.gov.sg/pioneers/pgp_overview.asp. Accessed 5 Nov 2014

Gray SF, Spry PG, Brookes ST, Peters TJ, Spencer IC, Baker IA, Sparrow JM, Easty DL (2000) The Bristol shared care glaucoma study: outcome at follow up at 2 years. Br J Ophthalmol 84:456–463

Green LV, Savin S, Lu Y (2013) Primary care physician shortages could be eliminated through use of teams, nonphysicians, and electronic communication. Health Aff (Millwood) 32:11–19

Hall TL, Mejia A (1978) Health manpower planning: principles, methods, issues. World Health Organization, Albany

Harper P, Kleinman E, Gallagher J, Knight V (2013) Cost-effective workforce planning: optimising the dental team skill-mix for England. J Enterp Inf Manag 26:91–108

Hawthorne L, Hamilton J (2010) International medical students and migration: the missing dimension in Australian workforce planning? Med J Aust 193:262

Homer JB (1996) Why we iterate: scientific modeling in theory and practice. Syst Dyn Rev 12:1–9

Homer JB, Hirsch GB (2006) System dynamics modeling for public health: background and opportunities. Am J Public Health 96:452–458

Hooker RS, Cawley JF, Everett CM (2011) Predictive modeling the physician assistant supply: 2010–2025. Public Health Rep (1974) 126:708–716

Huang CS, Cher T-L, Lin C-P, Wu K-M (2013) Projection of the dental workforce from 2011 to 2020, based on the actual workload of 6762 dentists in 2010 in Taiwan. J Formos Med Assoc = Taiwan yi zhi 112:527

Hulka BS, Wheat JR (1985) Patterns of utilization: the patient perspective. Med Care 23:438–460

Hurst K, Kelley Patterson D (2014) Health and social care workforce planning and development – an overview. Int J Health Care Qual Assur 27:562–572

Ishikawa T, Ohba H, Yokooka Y, Nakamura K, Ogasawara K (2013) Forecasting the absolute and relative shortage of physicians in Japan using a system dynamics model approach. Hum Resour Health 11:41–41

Kimball A, Resneck JS, Jacobson CC (2004) Generational differences in practice patterns of dermatologists in the United States. Arch Dermatol 140:1477

Kirch DG, Henderson MK, Dill MJ (2012) Physician workforce projections in an Era of health care reform. Annu Rev Med 63:435–445

Kurowski C, Wyss K, Abdulla S, Mills A (2007) Scaling up priority health interventions in Tanzania: the human resources challenge. Health Policy Plan 22:113–127

Lavieri MS, Puterman ML (2009) Optimizing nursing human resource planning in British Columbia. Health Care Manag Sci 12:119–128

Loh PT, Toh MP, Molina JA, Vathsala A (2015) Ethnic disparity in prevalence of diabetic kidney disease in an Asian primary healthcare cluster. Nephrology (Carlton) 20:216–223

Lomas J, Stoddart GL, Barer ML (1985) Supply projections as planning: a critical review of forecasting net physician requirements in Canada. Soc Sci Med 20:411–424

Machry RV, Tuchtenhagen S, Agostini BA, Da Silva Teixeira CR, Piovesan C, Mendes FM, Ardenghi TM (2013) Socioeconomic and psychosocial predictors of dental healthcare use among Brazilian preschool children. BMC Oral Health 13:60–60

Masnick K, Mcdonnell G (2010) A model linking clinical workforce skill mix planning to health and health care dynamics. Hum Resour Health 8:11–11

Maynard A (2006) Medical workforce planning: some forecasting challenges. Aust Econ Rev 39:323–329

Miller S (2012) The effect of the Massachusetts reform on health care utilization. Inquiry 49:317–326

North N, Hughes F (2012) A systems perspective on nursing productivity. J Health Organ Manag 26:192–214

O'connor PM, Alex Harper C, Brunton CL, Clews SJ, Haymes SA, Keeffe JE (2012) Shared care for chronic eye diseases: perspectives of ophthalmologists, optometrists and patients. Med J Aust 196:646

O'brien-Pallas L, Baumann A, Donner G, Murphy GT, Lochhaas-Gerlach J, Luba M (2001) Forecasting models for human resources in health care. J Adv Nurs 33:120–129

OECD (2013) Health at a glance 2013: OECD indicators. OECD Publishing, CANADA

Park E, Kim J (2015) Gender- and age-specific prevalence of metabolic syndrome among Korean adults: analysis of the fifth Korean National Health and Nutrition Examination Survey. J Cardiovasc Nurs 30:256–266

Park JM (2014) Health status and health services utilization in elderly Koreans. Int J Equity Health 13:73–73

Rhoads J, Ferguson LA, Langford CA (2006) Measuring nurse practitioner productivity. Dermatol Nurs 18(1):32–38

Roberfroid D, Leonard C, Stordeur S (2009) Physician supply forecast: better than peering in a crystal ball? Hum Resour Health 7:10–10

Scheffler RM, Liu JX, Kinfu Y, Dal Poz MR (2008) Forecasting the global shortage of physicians: an economic- and needs-based approach. Bull World Health Organ 86:516–523

Scott A (2011) Alternative approaches to health workforce planning. School of Population Health, University of Queensland, CANADA

Senese F, Tubertini P, Mazzocchetti A, Lodi A, Ruozi C, Grilli R (2015) Forecasting future needs and optimal allocation of medical residency positions: the Emilia-Romagna Region case study. Hum Resour Health 13:7

SINGHEALTH (2008) Singapore National Eye Centre opens new clinic in Balestier [Online]. Available: http://www.singhealth.com.sg/DoctorsAndHealthcareProfessionals/Medical-News/2008/Pages/Singapore-National-Eye-Centre-opens-new-clinic-in-Balestier.aspx. Accessed 5 Nov 2014

Spetz J, Trupin L, Bates T, Coffman JM (2015) Future demand for long-term care workers will be influenced by demographic and utilization changes. Health Aff (Project Hope) 34:936

Standard Chartered Bank (2014) Standard chartered launches Mobile Eye Clinic – Singapore's first comprehensive eye care on-the-move for senior citizens [Online]. Available: https://www.sc.com/sg/press-releases/en/_pdf/2014/Press_release_Standard%20Chartered_Mobile_Eye%20Clinic_27Sep2014_FINAL.pdf. Accessed 5 Nov 2014

Starfield B, Shi L, Grover A, Macinko J (2005) The effects of specialist supply on populations' health: assessing the evidence. Health Aff 24:W5

Sterman J (2000) Business dynamics: systems thinking and modeling for a complex world. Irwin/McGraw-Hill, Boston

Teo Z (2013) Educational profile of Singapore resident non-students, 2002–2012. In Department of Statistics Singapore, ed, OECD, FRANCE

Teye JK, Setrana MB, Acheampong AA (2015) Migration of health professionals from Ghana: trends, drivers, and emerging issues. Curr Polit Econ Afr 8:459

Thompson MJ, Christian Lynge D, Larson EH, Rosenblatt RA, Hart LG (2008) A longitudinal analysis of the general surgery workforce in the United States, 1981-2005. Arch Surg 143:345–350

Tjoa A, Kapihya M, Libetwa M, Schroder K, Scott C, Lee J, Mccarthy E (2010) Meeting human resources for health staffing goals by 2018: a quantitative analysis of policy options in Zambia. Hum Resour Health 8:15–15

Tomblin Murphy G, Mackenzie A, Guy-Walker J, Walker C (2014) Needs-based human resources for health planning in Jamaica: using simulation modelling to inform policy options for pharmacists in the public sector. Hum Resour Health 12:67–67

United States Bureau of Health Resources Development (1974) Registered nurses. In The supply of health manpower: 1970 profiles and projections to 1990

Van Der Heyden JHA, Demarest S, Tafforeau J, Van Oyen H (2003) Socio-economic differences in the utilisation of health services in Belgium. Health Policy 65:153–165

Wang J (2005) A review of operations research applications in workforce planning and potential modeling of military training. DTIC Document

Weiner JP (1994) Forecasting the effects of health reform on US physician workforce requirement: evidence from HMO staffing patterns. JAMA 272:222–230

Wendt C (2009) Mapping European healthcare systems: a comparative analysis of financing, service provision and access to healthcare. J Eur Soc Policy 19:432–445

Willis G, Woodward A , Cave S (2013) Robust workforce planning for the English medical workforce. In The 31st International Conference of the System Dynamics Society

Wong TY, Loon S-C, Saw S-M (2006) The epidemiology of age related eye diseases in Asia. Br J Ophthalmol 90:506–511

World Health Organization (2007) Task shifting to tackle health worker shortages. World Health Organization, Geneva

Wu M-H, Yu J-Y, Huang C-H (2013) Theoretical system dynamics modeling for Taiwan pediatric workforce in an era of national health insurance and low birth rates. Pediatr Neonatol 54:389–396

Yett DE, Drabek L, Intriligator MD, Kimbell LJ (1972) Health manpower planning: an econometric approach. Health Serv Res 7:134–147

Yett DE, Drabek L, Intriligator MD, KIMBELL LJ (1975) A microeconometric model of the health care system in the United States. Ann Econ Soc Meas 4:3

Zhang M, Katzman R, Salmon D, Jin H, Cai G, Wang Z, Qu G, Grant I, Yu E, Levy P, Klauber MR, Liu WT (1990) The prevalence of dementia and Alzheimer's disease in Shanghai, China: impact of age, gender, and education. Ann Neurol 27:428–437

Zheng Y, Cheng CY, Lamoureux EL, Chiang PP, Rahman Anuar A, Wang JJ, Mitchell P, Saw SM, Wong TY (2013) How much eye care services do Asian populations need? Projection from the Singapore epidemiology of eye disease (SEED) study. Invest Ophthalmol Vis Sci 54:2171–2177

Chapter 21
Lean Healthcare

S. Ali Torabi, Shirin Haddad Pour, and Nafiseh Shamsi G.

21.1 An Introduction to Lean Management

The concept of lean was first introduced and established in Toyota Company by Taiichi Ohno (1988), the Toyota executive. It is also known as Toyota's production system (TPS). The name "lean" was first introduced in the book "Machine that changed the world" by Womack et al. (1990).

The basis of the idea suggested by Taiichi Ohno was "do more with less". Hence, he asserted that by identifying and reducing/omitting wasteful activities in the company, it is possible to improve different aspects of performance. Therefore, he introduced the seven deadly types of wastes (or Muda in Japanese) as shown in Table 21.1. Muda can be defined as any activity that requires resources (e.g. man power, material, equipment, space, time); however, it does not create any value in return. On the other hand, value is defined by the ultimate customer and is recognized as a specific product (or a service) with a specific time and price at which the customer is willing to pay for it. The first step in establishing lean is to present a precise definition for what the value is in the eyes of the customers.

Hines et al. (2004) present the relationship between the value and cost of a product or service as viewed in Fig. 21.1

The dashed line is the equilibrium line between the cost and value in which the product cost is exactly the same as the value the customers perceive and are willing to pay for. When a product or service is located above this line it may be considered attractive for customers and as the distance from the equilibrium line grows when located above the line, the product or service is viewed more attractive\vadsjust{\pagebreak} to the customers. Based on the position of the

S.A. Torabi (✉) • S. Haddad Pour • N. Shamsi G.
School of Industrial Engineering, University of Tehran, Tehran, Iran
e-mail: satorabi@ut.ac.ir

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_21

Table 21.1 Seven types of wastes introduced by Ohno (1988)

Type of waste	Description
Over production	Producing the unnecessary product, or at the unnecessary time, or in the unnecessary quantity
Waiting (Delay)	Waiting for operations, resources, inspection, idle time, etc.
Transportation	Any type of transporting or the related activities (such as picking up, setting down, piling up, etc.) of the unnecessary items, related problems to transportation (may concern distance, utilization rate, or flow)
Inappropriate processing	Processes and operations performed without being absolutely necessary
Unnecessary inventory	Unnecessary retained inventory that may include material, parts, assembly parts (either as warehouse stock or in the form of work in process between processes)
Unnecessary motion	Unnecessary movements that do not add value such as searching for surgery tools during a surgery
Defects	Any costs related to the inspection, do-overs, repairs, and even customer complaints

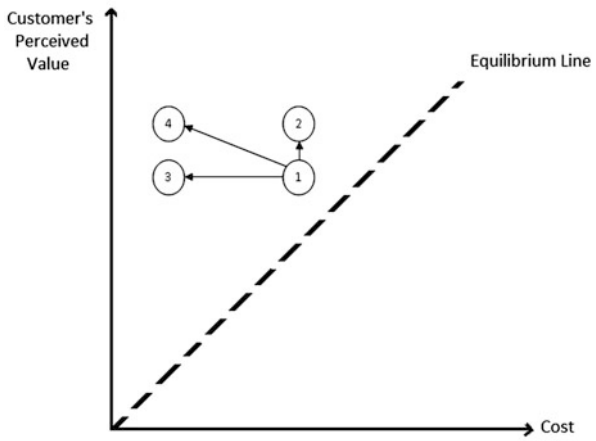

Fig. 21.1 Relation between the value, cost and waste (Adapted from Hines et al. (2004))

product, it can be concluded that an increase in the value of a product may occur in the form of one of the following cases:

Considering the first position, migration to position 3 happens when wasteful activities are eliminated and the associated costs are reduced, which in turn leads to increase in the overall value of the product. Migration to position 2 occurs by attempting to increase the value of product directly in the form of extra valuable features or services that do not cost anything for the company. For instance, smaller delivery batches or shorter delivery cycles may not incur additional cost for the company but create additional value for the customer. Migration to position 4 happens when increases in the aforementioned values happen at the same time.

21.1.1 Types of Activities

Lean management (LM) categorizes every work activity into the three classes:

21.1.1.1 Value Added (VA) Activities

These activities must satisfy at least one of the following conditions:

- It transforms information or material, or reduces the uncertainty level of the overall work.
- It is done correctly at the first time.
- It creates a sense of willingness to pay in the customer (either explicitly, or, in more complex systems, implicitly, which happens when the customer approves the activity and is willing to pay for it).

21.1.1.2 Required Non-value Added (RNVA) or Type I Non-value Added Activities

These activities do not meet the definition of value added activities, however; they cannot be eliminated since they are required. This requirement may be as the result of contract, company mandate, current technology, law, etc. Unavoidable movements of patients between wards and required medical tests are some examples of RNVAs.

21.1.1.3 Non-value Added (NVA) or Type II Non-value Added Activities

These activities consume resources (e.g. material and equipment), but do not lead to any added value. They are known as pure waste (i.e. Muda). For instance, unneeded emails or reports, and unnecessary medical tests are some examples of NVAs. By defining the value, even providing the right good or service in the wrong time is recognized as waste.

21.1.2 Lean Principles

Womack et al. (1990) introduced the first five lean principles in 'Lean Thinking' and the last principle was added to this set by Oppenheim (2011). Below, each principle is briefly explained.

21.1.2.1 Value

As we mentioned earlier, comprehending how the value is defined by the customers is a critical step in implementing lean management. Thus, the first principle is identifying the value desired by the customers.

21.1.2.2 Value Stream Mapping

This principle creates the value stream for each product or service. The value stream is the set of all the specific actions required to produce a specific product (either a good or a service, or both of them).

Ultimately, the value stream analysis almost shows that three types of actions are occurring along the value stream: (1) Some activities unambiguously create value, which are recognized as value added activities; (2) Some other activities create no value, yet they are unavoidable due to current definitions of the requirements, which are type I non-value added activities; And (3) those additional activities that do not create any value and have to be avoided immediately, which are non-value added activities.

The most important outcome of depicting a value stream map is to identify different variants of wastes when performing the different activities of the process under investigation in the form of type II non-value added activities, which should be removed and/or in the form of type I non-value added activities, which should be kept at the minimum amount.

21.1.2.3 Flow

This principle identifies the need to create a continuous flow for the product, or consequently, for the value. Hence, this principle aims to redefine the functions and any participants in the process, so they can make a positive contribution to create value. The best approach is to standardize processes based on the best performance to create a smooth flow of the product, which in turn allows the idle time to be used for innovation and creativity.

21.1.2.4 Pull Versus Push

This principle implements a ‘pull’ approach instead of ‘push’ between each two successive process steps, wherever it is not possible to create a continuous flow. It means that one activity requires to pay attention to the demand of the customer, that triggers events backwards through the value chain, before a shortage is about to happen. In other words, each process step is started if there is a request signal from its customer(s). Therefore, the customers are allowed to pull product or service from

the system anytime they require it, rather than the system pushing often unwanted products or services, onto the customer.

21.1.2.5 Perfection

After creating an environment in the process based on all the previous principles, this new principle concentrates on constantly aiming towards perfection. Accordingly, type I non-value adding activities are tried to be minimized and non-value adding activities are eliminated from the value stream.

21.1.2.6 Respect for People

Lean management recognizes that the people are the most important asset of any organization. Therefore, this principle overviews all the previous principles at all times. In a lean system as problems are introduced by people, the system is blamed not the messengers. Thus, no one is afraid to voice concerns or imperfections in the system. Key requirements for such a system are a culture of mutual respect and trust, open and honest communication, and synergistic and cooperating relationships of stakeholders (Oppenheim 2011).

21.2 Healthcare Wastes

As we mentioned in the previous section, waste terminology has been drawn from the manufacturing environment. Thus, to implement lean in healthcare it is imperative not only to translate the terminology but also to adopt it to healthcare environment appropriately. In this regard, many researchers have devoted their efforts to identify the healthcare wastes (Bowerman and Fillingham 2007; Radnor et al. 2012; Teichgräber and De Bucourt 2012; Fine et al. 2009).

21.2.1 Over Production

In a healthcare environment, this type of waste is more likely to come to view in the form of unnecessary tests or different types of unnecessary documents. Over production may be a result of fragmented or parallel care or simply an unorganized process. Each patient may be under the care of separate residents, attending, nurses and doctors. Therefore, duplicate chartings, multiple forms with the same information, or investigative 'just in case' tests may be ordered. Furthermore, a patient may undergo different departments in a healthcare system and an unorganized system also may result in filling forms with the same information more than once. Other

examples for over production may be providing copies of a form that is never used or automatically sending copies of reports or cc's on e-mails to those who have not asked for them and therefore will never use them. These all result in unnecessary filling of documents that later have to be organized.

21.2.2 Waiting (Delay)

Delays may have different causes such as: delivery of goods, occupied rooms, missing material or documents. Scheduling mistakes or busy schedules are some of main reasons leading to waiting. Patients, hospital staff, operating rooms, patient rooms, test results, prescriptions and medicines may suffer from delay. From as early as making an appointment to see a doctor over phone, patients may be delayed by waiting to be seen for an appointment or on hold for admission. Later, patients may be waiting for the process of diagnosis and treatment and their related delays, and finally waiting to be discharged by doctors. Hospital staff may have to wait for meetings, test results, office equipment (computer, photocopier, etc.), and other staff members. Rooms might be delayed due to scheduling mistakes, or necessary staff to discharge the current patient and then prepare the room, equipment, etc. for the next use. Test results may need analysis or staff members in order to test the specimens. Orders and prescriptions may suffer a time lag to be transferred between departments. Late arrival of any member of the staff as well as the patient may cause additional delays.

21.2.3 Transportation

Unnecessary movement of patients, staff members, documents and equipment may be considered as waste. Suffering from a poor layout or a poor management of equipment can be added to this type of waste. In other words, locating different departments of a hospital far from each other, or leaving items at a central space instead of locating them where they are frequently used, may create additional transportations. In this scenario, staff may need to walk to a different part of a ward or to a different department to attain documents, equipment, test results, charts, etc. Moreover, patients might be moved from one room at a building to another one in a different building in order to receive the necessary care due to remote locations (i.e. unsuitability of hospital's layout).

21.2.4 Inappropriate Processing

The first representative of this type of waste is stressed overworked staff. Ordering, delivering or implementing inaccurate procedures as well as excess material or personnel involved in the process are some types of inappropriate processes. Some information in a hospital may be subject to duplicity. For instance, patients may be asked to answer the same questions several times at different stages, which could be considered as redundant. Another form of redundancy is producing hard copies when a computer file is sufficient or the file has already been sent by an e-mail or posted via intranet. Some procedures at a hospital may need unnecessary pre-approval from someone that can create additional work for some staff.

21.2.5 Unnecessary Inventory

Any type of unneeded stocks and supplies are recognized as unnecessary inventory. This may come in the form of excess stock of equipment, medications, office supplies, surgical supplies, or any type of work in process that may exist in storage rooms or offices, or on units, or even hallways, many of which can be ordered on a just-in-time basis. Also, providing multiple locations for consumable goods that are not being used at a specific time is considered as unnecessary inventory. Excess inventory may prevent identifying shortfalls such as long setup times, long lead times, and late deliveries from suppliers.

21.2.6 Unnecessary Motion

This type of waste is defined as the excess movement of staff and information that are not required. Unnecessary motions of people are often related to ergonomic and layout issues. For instance, walking long distances to attain something is considered as waste. Moreover, unorganized placement of items such as placing heavier items on higher shelves and lighter items on the bottom shelves or placing those items (which are more frequently used than others) out of reach; causes excessive bending, reaching and walking to complete a process. In other words, not having the necessary items at the right places may create unnecessary movements. For instance, physicians and nurses may need to leave basic equipment or necessary information in the patient's room in order to limit avoidable movements.

21.2.7 Defects

Defects that may result in rework; represent great costs that may include associated costs of re-inspecting, rescheduling and capacity loss. Medical errors may vary from infections to mortality that incur a high emotional cost due to customer dissatisfaction as well as insurance costs or readmission costs and should be seriously avoided. Defects may be a result of incomplete information that may require repeating tests, placing the wrong labels on tubes or results, misreading the labels or handwritten orders, misfiling the documents, and sending out bills with an incorrect address.

21.3 Lean Management in Healthcare Systems

Trying to meet today's increasing patients' demands may cause lack of necessary resources in healthcare systems. Furthermore, medical supplies that were traditionally available on onsite storages are not available as before, due to realization of the storage and inventory wastes of overstocking. Additionally, logistics costs absorb a significant portion of the healthcare budget (Sinha and Kohnke 2009; Shah et al. 2008; Adebanjo et al. 2016). Therefore, while facing these challenges, continuous improvement approaches such as lean are really applicable in order to improve the efficiency and effectiveness of the healthcare systems (Wysocki 2004).

Public services, including healthcare, central government and local governmental organizations are amongst the service providers that embraced lean (Radnor 2010). From a historical point of view, lean was implemented in UK health service for the first time in 2001 and, later in 2002 in the USA (Radnor et al. 2012).

The work of D'Andreamatteo et al. (2015) shows an increment in the number of papers during the last 15 years that discuss implementation of lean management tools in the healthcare environment.

Brandao de Souza (2009) discusses that lean healthcare in the private sector has proven to be effective and it is evident that most of the work done on lean healthcare has been carried out in the USA within the private healthcare systems. Additionally, a significant increase in the number of public healthcare systems implementing lean tools such as in the UK shows that a need for lean thinking has been identified in those sectors as well.

The lean principles which were generally discussed in the first section are applicable in every environment such as healthcare systems as well as manufacturing systems but evidently in different ways. The first step in order to start becoming lean is to identify value. Therefore, it is of foremost importance that we first recognize customers and their needs. Identifying the customers' needs at a healthcare system can be quite challenging. Not only patients, but also clinicians and insurance companies especially in public sectors, are in the customer group (Kollberg et al. 2006). However, patients are identified as the prime customers

for healthcare environment, since a healthcare system is established to provide necessary services to them. Thus, value is mostly identified with regards to the patients' views (Wickramasinghe et al. 2013; De Souza and Pidd 2011; Kollberg et al. 2006; Dickson et al. 2009). In the patient-centered view, qualified care that is efficient, appropriate, timely and safe must be provided by the healthcare system in order to increase the level of patients' satisfaction (Tolga Taner et al. 2007).

As we discussed in Sect. 21.1, in order to recognize waste along the whole value chain, the first and the foremost important tool is using the value stream mapping to depict the whole process in a visual map by which the non-value added activities could be distinguished from the rest through an exact leanness analysis (i.e. using lean tools such as five-why analysis). In this way, the system analyst can focus on eliminating these activities in order to follow the lean thinking and make the value chain as lean as possible. By using VSM for the entire process from making the first contact by the patient to his/her discharge, all the activities are to be categorized into the three groups, namely, value-added activities, required non-value added activities, and non-value added activities for which a detailed explanation was presented in Sect. 21.2. Bowerman and Fillingham (2007) assert that typically, in a day-to-day life at a hospital, nine activities out of ten are not value adding activities, which may be categorized into the delays, interruptions, errors, and unnecessary movements. The rest of the principles are to follow as explained in Sect. 21.1 for such activities.

21.3.1 Lean Tools and Techniques

There are several tools and techniques that can help healthcare system in order to implement lean thinking. Here, we are going to explain some of the main ones. Among these tools and techniques, VSM is the starting point to identify applicable lean tools in each case. As shown in Fig. 21.2, VSM organizes the flow of activities. However; it can also include other important data such as information flow, cycle time, batch size, and so on. Therefore, by providing a big picture from the process under investigation, VSM offers a common language for everyone involved. Furthermore, by understanding the relationship between the patient flow and the information flow, lean techniques and concepts can be implemented to make informed decisions using the operations research (OR) techniques (Jimmerson 2009).2

21.3.1.1 Root-Cause Analysis

By using this technique, potential (or happened) errors in healthcare-related processes are examined and the root-cause(s) of each error is (are) studied. As the result, suitable countermeasures are taken to prevent them from further happening (Wysocki 2004). Simons et al. (2014) utilized the major categories in this diagram, namely people, machines, methods and materials in order to reach the goal of having

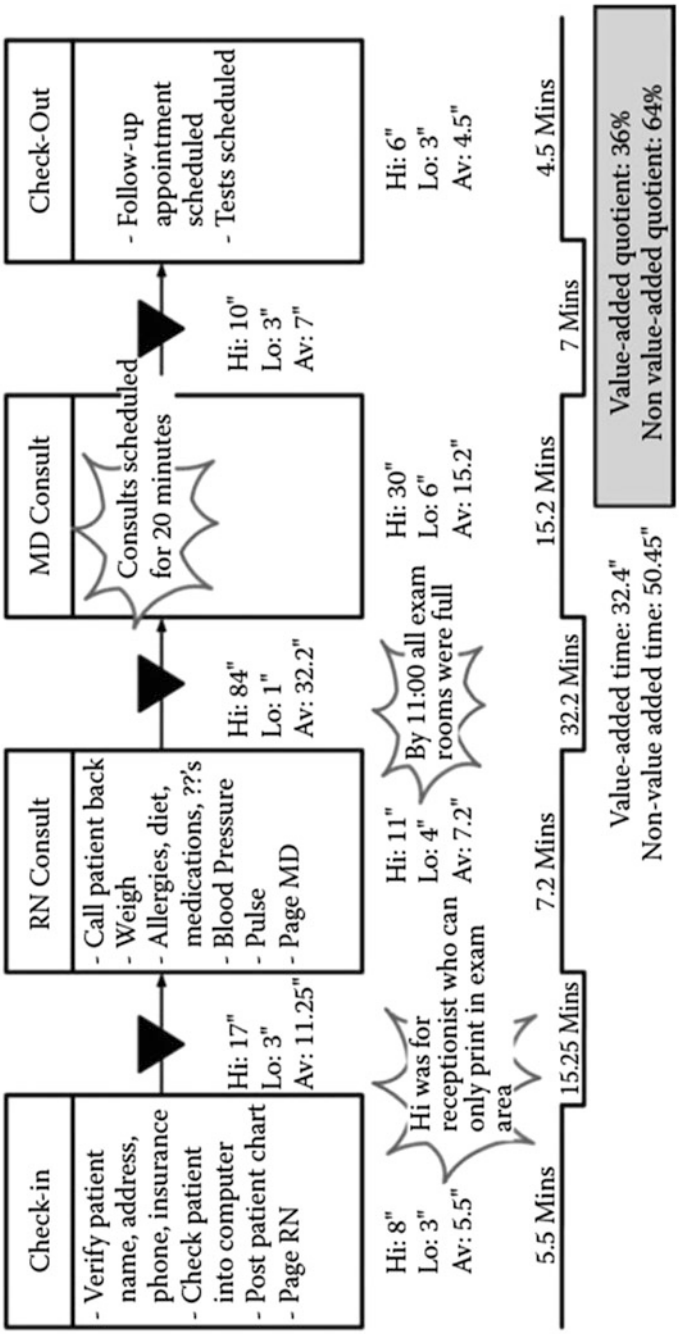

Fig. 21.2 An example of value stream mapping (Jimmerson 2009)

zero door movement for the operation room door during surgery and found 13 root causes for this problem.

21.3.1.2 Batch Size Reduction

Minimizing the batch size in healthcare-related processes, resulting in the ultimately single-piece flow, leads to minimizing or eliminating work-in-process, and decreasing the turnaround (i.e. flow) time. This technique may be applicable in a healthcare system where instead of visiting three patients at a time before sending their test samples to the lab, the nurse or physician in charge sends each sample to the lab as soon as she/he sees each patient in order to minimize pile ups that may happen during the wait for one test to be completed before running the other two tests (Villa 2010).

21.3.1.3 The A3 Problem-Solving Report

Suitable for healthcare environments, this report documents the essential stages of the problem which is faced during the day-to-day processes to find an appropriate solution (Sobek and Jimmerson 2006).

Jimmerson et al. (2005) have used the A3 tool to identify dozens of problems in a pathology lab (Fig. 21.3). For instance, the paperwork associated with specimens from different stages and the specimens themselves did not align at times. This resulted in wasteful hours that specialists had to manually match them together. The countermeasure devised for this problem was using specific software by which the related information could be organized and printed in the same order as the flow of the specimens.

21.3.1.4 Takt Time

Takt time is defined as the amount of available time in which a step of one process should be done in order to satisfy the customer's demand. In order to reduce the waste in the process, takt time has to be minimized (Sloan et al. 2014b). Similar processes that are constantly performed in a healthcare environment such as routine check-ups, making rounds, preoperative or postoperative procedures for some surgeries and so on; may become standardized in order to minimize waste and maximize the efficiency of the system.

21.3.1.5 Kaizen Blitz (Kaizen Event)

A Kaizen event or kaizen blitz involves small steps to improve the performance of a process continuously with a limited time line (Doolen et al. 2008). For example,

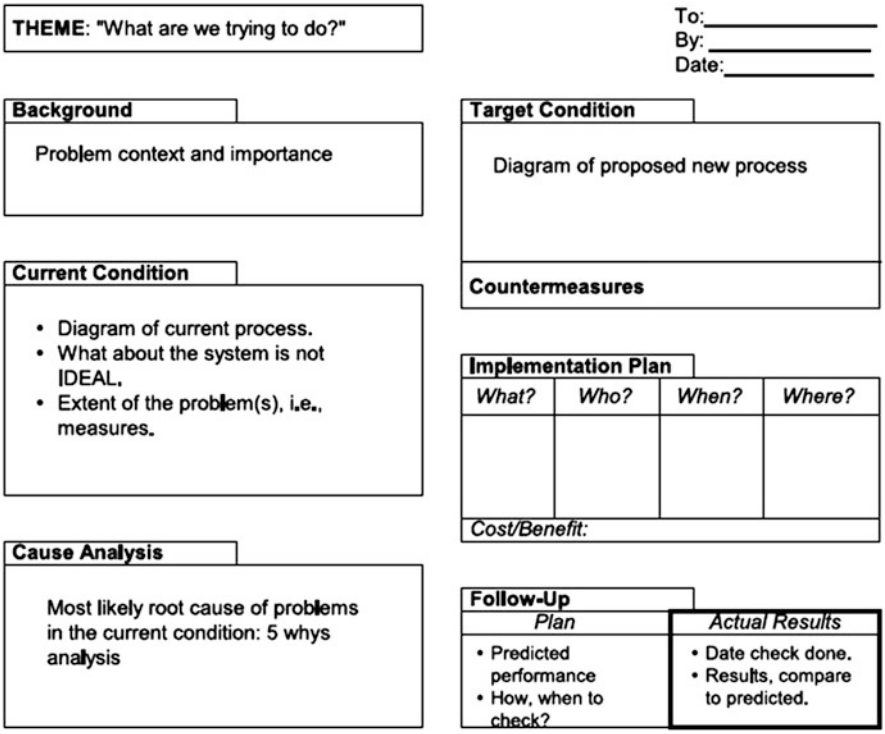

Fig. 21.3 An A3 template (Jimmerson et al. 2005)

in order to decrease late discharges in a hospital, many smaller objectives can be performed daily by the staff in order to reach that goal (e.g. decreasing delays, mistakes, unnecessary motions and so on).

21.3.1.6 Quick Changeover

This technique, which originates from manufacturing setting where it is also called *Single-Minute Exchange of Die* (*SMED*), may be applicable in healthcare setting as well. This lean technique aims to minimize the required set-up times through eliminating unnecessary activities and/or doing some activities before starting the set-up operations (Villa 2010). An example is to provide the necessary equipment/tools/material available for a specific process (e.g. surgery operations) that is common in a ward in order to decrease the required time that the staffs take for gathering such requirements.

21.3.1.7 Total Productive Maintenance (TPM)

By involving all the employees in the system, TPM uses a preventive maintenance system for equipment (especially those critical equipment such as radiology and MRI equipment) in order to improve their reliability through minimizing their down times and improving their working conditions (Singh and Singh 2009).

21.3.1.8 Kanban (Pull)

Kanban is an inventory control system to control material flow in a logistics chain. In a specially designed box/container, a Kanban card acts as a signal, which is sent to the supplier or warehouse indicating that a work station's inventory is low on a specific item and an order is being placed in the work station. Kanban system aims to reduce the work in process inventory. Other than items that use Kanban for the "just in time ordering", this method may be used to create signals to pull patients in the healthcare system when it is possible. In other words, as soon as a patient is scheduled to be discharged, a new patient should be scheduled to take the empty bed (Bowerman and Fillingham 2007).

21.3.1.9 Industrial Housekeeping (5S)

The following steps are the principles of the 5S:

- *Seiri* (sort): Classify items into groups of required and not required items.
- *Seiton* (straighten): A systematic arrangement for everything.
- *Seiso* (shine): Wash and clean everything in the work place.
- *Seiketsu* (standardize): Analyze the best procedures to form routines.
- *Shitsuke* (sustain): Regulate to maintain the improvements.

Noteworthy, 'safety' and 'security' have also been recommended to be added as additional Ss to aforementioned 5S principles (Bowerman and Fillingham 2007).

This method creates a safe, organized, and clean way of doing everyday procedures in which problems could be easily visible. For instance, an organized nursing station which has closet including labeled equipment at appropriate places reduces the mistakes associated with not finding the required equipment at the right time that they are needed. As time becomes an extremely important factor, especially in a trauma stabilization unit, the necessity of performing 5S increases.

21.3.1.10 The Five Whys Technique

In order to create an effective workplace and uncover wastes, this tool asks 'why' several times until the root cause of a problem is identified. Usually, the answer lies after five whys; therefore, this simple method is called the five whys. For instance,

consider the case in which a nurse was not found at her station. By asking why she was not at the station, it was discovered that by a patient's complaint the nurse had to go to the supply room to fill an order that the supply room had neglected. By continuing to ask why, it was resolved that the stock of the equipment required was depleted and that the supply room lacked a system for informing them of any equipment that they run out of, which later was replaced by implementing Kanban to pull required equipment just-in-time (Zidel 2006).

21.3.1.11 Poka-Yoke/Jidoka (Mistake-Proofing)

This is a mechanism that requires the availability of necessary conditions before trying to execute a process step, in order to prevent any potential failure during the process. Where it is not possible to prevent failures from happening, Poka-Yoke tool discovers and eliminates them. In order to insure reliability, the process is stopped when a failure is discovered (Singh and Singh 2009). As an example, since appropriate diagnosis and further treatments are based on the precision of hemoglobin testing, an automatic hemoglobin testing device that performs the analysis under a minute without relying on visual judgment or complicated methods can mistake-proof this test.

21.3.1.12 Standardized Work

This method properly structures all the steps of a process to increase its performance and eliminate potential wastes. The aim of work standardization is to reduce variation and complexity when doing any process/activity. In a healthcare system, in addition to admission or check out procedures or paperwork, some other procedures may be standardized to improve the efficiency and decrease errors such as rehabilitation mistakes (Bowerman and Fillingham 2007).

21.3.1.13 Visual Controls

This method displays the necessary information which is available accurately, completely, and noticeably. Visual controls have four levels: A visual indicator only aims to transfer specific information, for instance, placing a sign at the door where special dietary instructions or treatments are to be followed for a patient. Second is a visual signal that creates an alert or alarm. For example, if patients require help, they may use a switch in their rooms that is connected to a nurse's call light. Third, to direct actions, visual controls may be used. An example may be a closed position in full containers in order to avoid overflowing. Fourth level is foolproof and only permits a specific action. For instance, a Bloodloc Safety System is used to insure that the properly cross-matched unit of blood is transfused to the right recipient (Zidel 2006).

Even though these tools and techniques are important in order to implement lean in any system, they are not enough to establish and sustain a lean system. In order to fully commit to a lean healthcare, preparedness or readiness factors must be considered (Rees 2014; Sloan et al. 2014a; Radnor et al. 2006; Holden 2011). Sloan et al. (2014a) offer a comprehensive outline of readiness factors by reviewing relative material. For example, not only management understanding, dedication and support for lean, but also creating an environment based on trust and mutual respect for getting people's ideas; help the processes to be implemented in a healthcare system. This also helps cultivate lean culture among staff which is essential for lean to work. In order to involve and encourage the staff to implement lean, continuous training sessions about lean concepts/tools are essential. Another technique to encourage staff to properly communicate and implement lean tools is to have an adequate bonus plan to create the necessary incentives. Additionally, lean must be viewed as a long-term strategy when it comes to the healthcare setting's strategic agenda. Furthermore, in order to clarify objectives and needs of each group, any possible customer group must be identified and each customer's needs are met accordingly to create value. This helps to avoid resistance to change in the system as well as dissatisfied customers and staff. As mentioned before, creating a complete and accurate VSM has an important role to identify the current wastes and recognize the necessary tools required to reduce/remove identified wastes. Therefore, going through this step properly is also recognized as a readiness factor (Sloan et al. 2014a). There exist three well recognized, significant and most referred lean healthcare systems in the literature, namely, Virginia Mason Medical Centre, Seattle, USA, Royal Bolton Hospital, Bolton UK and Flinders Hospital, Adelaide, Australia, which have established such readiness factors in their systems and acknowledge their importance (Rees 2014).

Additionally, by reviewing articles from 2000 to 2012, Andersen et al. (2014) introduce the most important facilitators for implementing lean in healthcare systems as management, supportive culture, training, teamwork and accurate data.

21.3.2 Lean System from Employees' Perspectives

Employees play a crucial role in the process of creating a lean system. Therefore, it is of foremost importance to view how implementing lean would affect their work.

21.3.2.1 Task Identity

One of the principles of lean is the flow of the value without interruptions or queues, without batches and with no inventory or NVA activities. To fulfill this goal, employees may be required to do several tasks continuously to aid the flow of the value. Employees who are trained to be multi-skilled can better recognize their task identity, because, they would have a better grasp of different aspects of

a work and how they come together to create the final outcome (De Treville and Antonakis 2006).

21.3.2.2 Feedback

The third principle of lean, flow, creates a better environment for feedback to travel upstream in the system (De Treville and Antonakis 2006). Whereas, providing a smooth and instant feedback to downstream employees in the system requires the aid of visual controls (Mann 2014).

21.3.2.3 Work Facilitation

To facilitate the work for employees, several lean tools that were described earlier, such as visual controls, Poka-Yoke, and 5S may be found effective (Sloan et al. 2014b).

21.3.2.4 Responsible Autonomy

In a lean system, teamwork is recognized as the primary work unit and responsible for performing daily tasks. Mutual trust within teams and between employees and managers is also required for lean implementation. Self-managed teams that all work towards making a consensus of decision; help with devolution and sharing of the power within the organization. It helps increasing responsible autonomy (Sloan et al. 2014b).

21.3.2.5 Choice Autonomy

There are two opposing effects of the lean implementation. On one hand, individual and team involvement in order to understand what the problems are and make informed decisions increases accountability and self-sufficiency between employees. On the other hand, in order to limit waste and decrease the necessary resources, and also in the presence of leans tools such as standardized work, and Takt time; a decrease in choice autonomy is evident which in turn causes an increase in job depression (Sloan et al. 2014b).

21.4 Strategic and Operational Challenges in Hospitals

Those who implement lean in healthcare systems may experience several ongoing challenges. Some of these challenges are discussed in this section.

21.4.1 Organizational Challenges

Healthcare systems are recognized as highly political and sophisticated organizations run by powerful experts. Thus, any improvement or management technique cannot easily put to work in healthcare environments in comparison to other organizations (Pettigrew et al. 1992; Radnor et al. 2012).

21.4.2 Defining Waste

In the healthcare environment, due to significant differences between patients' cases and scenarios, it is not always easy to define value added activities. Moreover, since each individual has its own method of finishing a task, the definition of waste is not always the same when it comes to different people's perspectives (Grove et al. 2010b).

21.4.3 Cultural Barriers

An important challenge in front of the hospitals is a cultural clash due to the fact that lean is originated from the manufacturing environment (i.e. Toyota production system). Therefore, many people in a healthcare system challenge if this principle stands in an environment that if a mistake is made, lives may be lost. Their view disregards lean because they presume that a culture of efficiency does not have a say when it comes to caring for patients' lives. In order to lower the effects of this culture clash, it has been suggested that lean should not be called Toyota Production techniques, and only to be referred as lean techniques in a healthcare environment (De Souza and Pidd 2011; Wysocki 2004).

21.4.4 Fear of Losing Jobs

When it comes to lean activities, especially when it involves a team or organization with little experience, one of the biggest fears is that it is intended or will lead to job cuts. This cynicism may create an environment in which the staff does not communicate honestly and completely about problems or concerns (Fine et al. 2009).

21.4.5 Slow Engagement of the Physicians

Even though physicians are recognized as the process owners, they are often slow to involve in accepting lean, because, they fear that lean principles aim to prioritize productivity and efficiency over quality and patient experience. Nevertheless, if they participate in its implementation, usually, they are able to make valuable contributions and tend to involve other staff members (Waring and Bishop 2010; Wu et al. 2010).

21.4.6 Sustaining Lean System

The important part of implementing lean principles is to make sure that they will be applied in time. Therefore, remaining lean after the initiatives are established in the organization and view lean as a long-term commitment, is an important challenge in front of a lean healthcare system (Fine et al. 2009).

21.4.7 Negative Connotations

Negative images of trimming fat, shredding jobs, or targeting ruthless efficiency are some of the negative connotations associated with lean implementation that may keep the frontline staff suspicious and resistant towards lean (Wu et al. 2010).

21.4.8 Defining the Customer

Even though the patient was identified as the end customer, there may be scenarios that consider other customer groups such as, government agencies, insurance companies, charities and families of the patients. At times the value defined by these groups may appear contradictory (Grove et al. 2010b).

21.4.9 Measuring Value

In healthcare systems, the price of product or service is seldom tangible or realized by the patients, and the quality of the service cannot be fully quantified. Therefore, it is more difficult to estimate the value in healthcare environment. However, not only for patients, but also for internal purposes, it is vital to suggest objective indicators for calculating the value (Dickson et al. 2009).

Managers may easily overcome some of these challenges by building trust, providing incentives, hospital-wide communication efforts such as hospital newsletters or posters, correctly handling employee feelings, involving employees, open channels of communication, and other readiness factors discussed earlier.

21.5 Using Operations Research Tools for Lean Healthcare

In addition to concepts and principles discussed earlier, Operations Research (OR) techniques mainly mathematical programming approaches (e.g. mixed integer programming) and other OR tools (e.g. simulation optimization and queuing theory) would help to reduce healthcare systems' wastes/costs. Accordingly, OR scholars are increasingly developing OR-related (i.e. analytical/quantitative) tools (e.g. mathematical and simulation models) to improve the performance of different processes within the healthcare systems (Bhattacharjee and Ray 2014; Harper et al. 2010; Hong and Yaoqiu 2010; Paul and Jotshi 2013).

There are several issues for which analytical tools (e.g. simulation and optimization models) can be applied to reduce the wastes of healthcare systems. This may include: (1) developing the inventory control models to minimize shortage or excess inventory of medical perishable goods, (2) maintenance scheduling of medical equipment to minimize their down times, (3) scheduling of operation rooms and shift changing processes to maximize patients', physicians' and nurses' satisfaction levels while meeting resource constraints, (4) bed capacity management to maximize their utilization rate, and (5) risk assessment and management to identify potential risks in the healthcare-related processes and responding to those threats with high risk levels through preparing suitable response plans (see for instance (Torabi et al. 2016) for more details in the risk assessment process). Therefore, quantitative modeling approaches (e.g. mathematical models) should be taken into more consideration in parallel to traditional lean tools such as VSM and 5S to reach ultimate lean thinking goals (i.e. Quality, Cost and Delivery, which are briefly called QCD objectives in the context of lean management). Hereafter, we provide some exemplars of OR tools' applications in dealing with healthcare-related decision problems in order to elaborate on the application of OR techniques towards reaching to lean health systems.

In the healthcare services, scheduling is considered as a complex subject because various aspects of a healthcare system such as resources, patients' requirements, the culture and structure of the organization, and medical perception and criteria have to be managed, where some of which may introduce new constraints to the model. There are also some other factors that may influence the scheduling problem (e.g. any specific timetable that a physician may introduce, the level of emergency that a patient's condition may have either real or perceived, or any additional constraints that a patient's insurance may enforce). The objectives of these models are often pragmatic, for instance, minimizing cost, optimizing room availability, or regarding practitioner or staff. While many of these components are assessable, some other

confounding factors may not be as easily measured. For example, struggling with prolonged waiting times may cause an increase in costs, patient, practitioner or staff irritation, and operational inadequacies. The irritation factor may not always be as tangible as other factors but just as important. Due to the importance of this area of research, many papers have devoted their attention to scheduling problem (Beliën and Demeulemeester 2007; Blake and Donald 2002; Van Oostrum et al. 2008; Simon and Canacari 2014).

Another category is the process improvement. By creating a one-stop pre-operative clinic before the surgery, van Vliet et al. (2010) and Siddique et al. (2012) were able to decrease the mean waiting time for surgery and the mean physician time for each patient in their models. Another optimization model was developed by Waldhausen et al. (2010) for a pediatric surgery clinic that resulted in an increase in the number of patients treated in 4 h periods, an increase in the efficiency of the physicians in the time they are allocated to each patient, a decrease in the exam room time for each patient, and finally better overall survey results considering patient's experience in the clinic.

Moreover, introducing models to improve the efficiency of operating rooms is another example in this area. Schwarz et al. (2011) provides statistical analysis for a model that aims to reduce the turnaround time (i.e. the time between each two cases that the physician is not treating a patient or in this case operating), and the throughput time- the duration of time between the entry and the exit of the patient which in this paper is the time a patient is in the operating room from entry to exit. Collar et al. (2012) also achieved to reduce the turnover time- the duration of time from the release of one patient to the entry of the next, as well as the turnaround time. Whereas Adams et al. (2004) were able to reduce the turnaround time by 32–36% and the turnover time by 30%, Niemeijer et al. (2013) introduced a model in hip fracture surgeries that resulted in a 36-percent decrease in the duration of these surgeries.

Another important area is analyzing the risks and minimizing the related errors. In the work of Yousri et al. (2011), a model was introduced with an effort to reduce the mortality rate and succeeded by bringing the overall mortality rate from 21% to 11% and the 30-day mortality rate from 12% to 7% in patients with fractured neck of femur.

In order to manage costs in a hospital, Gayed et al. (2013) reported that it is possible to have a one-million-dollar annual saving by the means of appointing one additional work force and decreasing the amount of out-sourcing to other organizations. Other examples can be viewed in Collar et al. (2012), Adams et al. (2004), and Niemeijer et al. (2013).

Many have suggested methods to reduce the length of stay of patients in the hospital. In this regard, Iannettoni et al. (2011) and Niemeijer et al. (2010) showed reductions of 64% and 28% respectively. Niemeijer et al. (2013) presented a 31-percent decrease, whereas Gayed et al. (2013) reported a 36-percent decrease.

21.6 Case Studies

In this section, by reviewing six case studies in the literature, an overall evaluation of lean healthcare systems in different scenarios is presented.

21.6.1 ThedaCare, Northeastern Wisconsin, USA

Toussaint (2009) introduces the effects of lean system in two acute care and two community hospitals in northeastern Wisconsin, that all are under the same healthcare system, ThedaCare. With 5300 employees and 20,868 annual patient admissions, ThedaCare is considered as a major care provider in Wisconsin. Therefore, with the help of lean, they were able to eliminate millions of dollars of waste. This system established lean as a long-term plan with five weekly Kaizen blitz to improve care while lowering the costs. Some of the results of lean implementation for instance are, better heart attack response rates, a decrease in delivering preterm births, offering same-day appointments in every clinic and office, etc. In a course of 3 years ThedaCare has managed to attain the first rank for the lowest costs in the state and reduce the hospital costs by an amount equal to 5% of the yearly revenue. Moreover, from 2006 to 2009 that the research was published, a 12-percent increase in productivity with the help of lean, resulted in $27 million saving for ThedaCare, thus, it was able to keep price increase rate half the rate of other competitors.

Another effect of lean with kaizen blitz was the decrease in the mortality rate for coronary bypass surgery, from nearly 4% in 2002, to 1.4% in 2008 and 0% for half of 2009. Also, the average inpatient stay for this type of surgery dropped from 6.3 to 4.9 days as well as, a 22-percent decrease in costs.

21.6.2 F. Miulli Hospital, Italy

Lasorsa et al. (2015) presented a case in operating room of the Italian Hospital, F. Miulli. The storage areas were organized based on the 5S and Visual Controls with an estimated saving of 13,000 € as the result of time decrease in searching for required tools. Also, an additional estimated saving of 229,421 € was due to further improvements in inventory that created more availability with the means of Batch Size Reduction of tools that were sent to the Central Sterile Supply Department to be processed. Other benefits from applying lean management in F. Miulli may include, a 33-percent decrease in the cycle time of preparing surgical tools, a 70-percent increase in the utilization of the tools during the preparation of surgical case cart, a 47-percent decrease in nonconformities during the preparation of surgical case

cart, and a 75-percent decrease in nonconformities after the preparation of surgical case cart.

21.6.3 Flinders Medical Centre, Adelaide, Australia

Ben-Tovim et al. (2007) provide the case study of Flinders Medical Centre which is a primary regional care provider with busy Emergency Department typically seeing 50,000 patients annually, 40% of whom need further care and are therefore admitted. Implementation of lean principles resulted in a 14-percent decrease in the average time patients spent in Emergency Department. Furthermore, creating a less chaotic environment in Emergency Department, the number of patients who did not complete their care dropped to 3% from an earlier 7%. Also, the average inpatient stay was reduced from 5.7 to 5 h in the first year, and although a 10-percent increase in the number of patients occurred in the next year, inpatient stay saw another 6 min drop. Moreover, the hospital was able to reduce the number of adverse cases that needed the interference of insurers by half.

21.6.4 A Primary Care Trust of National Health Service, UK

Grove et al. (2010a) studied the effects of lean implementation in a health visiting service in a Primary Care Trust (PCT) with 2400 employees and 217,000 patients annually for a period of 13 months.

With the help of a comprehensive VSM, 67 processes were identified in the health visiting service, 65% of which was later recognized as waste that needed to be eliminated or redesigned. Clinical staff spent a range of 12–70% of their time occupied with VA activities and a range of 24–63% of time was spent on NVA activities that had to be eliminated altogether. It was realized that a large amount of waste did not require any time-consuming or costly organizational modifications and was easily eliminated with defining standardized work or carrying out a 5S project.

21.6.5 Royal Bolton Hospital, Bolton, United Kingdom

Bowerman and Fillingham (2007) conducted an 18-month study on Royal Bolton Hospital, which is now considered as an informative reference to the application of lean in healthcare systems. They used lean principles and took advantage of tools such as 5S, standardized work, visual controls, etc. As a result, paper work has been reduced by 42%, inpatient stay has been shortened by 33%, and mortality rate of patients decreased by a third. Moreover, the time for necessary blood processes was

down to 2 h from an earlier 2 days, and the average time to get results ready in pathology was decreased significantly from 24 to 2–3 h. 3.1 million pounds was the financial benefit saved directly by lean implementation in Royal Bolton Hospital.

21.6.6 Three Hospital Sites, New Zealand

Rees (2014) provided three hospital sites with differences in organizational readiness when it comes to implementing lean techniques. These three sites all used VSM, 5S, and Standardized Work as their lean tools. Site A with 37,000 patients that yearly come to Emergency Department of whom 25–30% are admitted to the hospital, Site B has 40,000 annual visits with 23% admission rate, and site C that has a 28,000 patients visiting the Emergency department and 30% require further care. Top management at sites A and C were not involved in lean implementation; However, site B's top management level was found to be engaged with implementing a lean system in the entire hospital. At site A, the time for some of the processes required for a diagnosis, were reduced almost by half. Site B and site C were able to reduce the admission time for the patients. Also, site C reduced errors and reworks.

21.7 Conclusion

Applying Lean Management techniques in healthcare systems can move them from their traditional task orientation to patient flow orientation, which in turn can provide better patient service, patient care, and utilization of assets.

There are lots of wastes, which occur in healthcare systems and degrade their performance (e.g. efficiency, effectiveness and productivity). The implementation of lean tools such as value stream mapping (i.e. VSM), Kanban inventory control mechanism and industrial housekeeping (i.e. 5S) in a healthcare setting, particularly in a hospital, would remove duplicate processes and unnecessary procedures such as recording patient details in multiple formats and places; patients being moved towards before beds are available; patients being moved from one ward to another; excessive waiting for doctors and consultants; and uncoordinated, variable discharge processes resulting in a longer length of stay than necessary.

D'Andreamatteo et al. (2015) asserted that there are several unexplored issues in the lean healthcare environment that need further attention. They mentioned that the advantages and disadvantages of combining lean with other methods; requires a more comprehensive research. Currently, most of the attention of the researchers is focused on intensive care or emergency rooms; however, it is necessary to investigate lean in other settings such as primary care, home care, or other departments or areas in healthcare systems. Furthermore, studying lean healthcare in different countries provides a deeper understanding of the possible

effects that cultural behaviors might have on lean management. And finally, studies regarding the cost efficiency of lean interventions need deeper investigation.

References

Adams R, Warner P, Hubbard B, Goulding T (2004) Decreasing turnaround time between general surgery cases. JONA 34:140–148

Adebanjo D, Laosirihongthong T, Samaranayake P (2016) Prioritizing lean supply chain management initiatives in healthcare service operations: a fuzzy AHP approach. Prod Plan Control 27(12):953–966

Andersen H, Røvik KA, Ingebrigtsen T (2014) Lean thinking in hospitals: is there a cure for the absence of evidence? A systematic review of reviews. BMJ Open 4:e003873

Beliën J, Demeulemeester E (2007) Building cyclic master surgery schedules with leveled resulting bed occupancy. Eur J Oper Res 176:1185–1204

Ben-Tovim DI, Bassham JE, Bolch D, Martin MA, Dougherty M, Szwarcbord M (2007) Lean thinking across a hospital: redesigning care at the Flinders Medical Centre. Aust Health Rev 31:10–15

Bhattacharjee P, Ray PK (2014) Modelling and optimisation of healthcare systems: an overview. Int J Logist Syst Manag 19:356–371

Blake JT, Donald J (2002) Mount Sinai hospital uses integer programming to allocate operating room time. Interfaces 32:63–73

Bowerman J, Fillingham D (2007) Can lean save lives? Leadersh Health Serv 20:231–241

Brandao De Souza L (2009) Trends and approaches in lean healthcare. Leadersh Health Serv 22:121–139

Collar RM, Shuman AG, Feiner S, Mcgonegal AK, Heidel N, Duck M, Mclean SA, Billi JE, Healy DW, Bradford CR (2012) Lean management in academic surgery. J Am Coll Surg 214:928–936

D'Andreamatteo A, Ianni L, Lega F, Sargiacomo M (2015) Lean in healthcare: a comprehensive review. Health Policy 119:1197–1209

De Souza LB, Pidd M (2011) Exploring the barriers to lean health care implementation. Public Money Manag 31:59–66

De Treville S, Antonakis J (2006) Could lean production job design be intrinsically motivating? Contextual, configurational, and levels-of-analysis issues. J Oper Manag 24:99–123

Dickson EW, Singh S, Cheung DS, Wyatt CC, Nugent AS (2009) Application of lean manufacturing techniques in the emergency department. J Emerg Med 37:177–182

Doolen TL, Van Aken EM, Farris JA, Worley JM, Huwe J (2008) Kaizen events and organizational performance: a field study. Int J Product Perform Manag 57:637–658

Fine BA, Golden B, Hannam R, Morra D (2009) Leading lean: a Canadian healthcare leader's guide. Healthc Q 12:32–41

Gayed B, Black S, Daggy J, Munshi IA (2013) Redesigning a joint replacement program using Lean Six Sigma in a veterans affairs hospital. JAMA Surg 148:1050–1056

Grove AL, Meredith JO, Macintyre M, Angelis J, Neailey K (2010a) Lean implementation in primary care health visiting services in National Health Service UK. Qual Saf Health Care 19:e43–e43

Grove AL, Meredith JO, Macintyre M, Angelis J, Neailey K (2010b) UK health visiting: challenges faced during lean implementation. Leadersh Health Serv 23:204–218

Harper PR, Powell N, Williams JE (2010) Modelling the size and skill-mix of hospital nursing teams. J Oper Res Soc 61:768–779

Hines P, Holweg M, Rich N (2004) Learning to evolve: a review of contemporary lean thinking. Int J Oper Prod Manag 24:994–1011

Holden RJ (2011) Lean thinking in emergency departments: a critical review. Ann Emerg Med 57:265–278
Hong L, Yaoqiu W (2010) An optimized strategy based on demanders' satisfaction for rescue resource allocation. In Logistics systems and intelligent management, 2010 international conference on, 2010. IEEE, p 531–534
Iannettoni MD, Lynch WR, Parekh KR, Mclaughlin KA (2011) Kaizen method for esophagectomy patients: improved quality control, outcomes, and decreased costs. Ann Thorac Surg 91:1011–1018
Jimmerson C (2009) Value stream mapping for healthcare made easy. CRC Press, Boca Raton
Jimmerson C, Weber D, Sobek DK (2005) Reducing waste and errors: piloting lean principles at Intermountain Healthcare. Jt Comm J Qual Patient Saf 31:249–257
Kollberg B, Dahlgaard JJ, Brehmer P-O (2006) Measuring lean initiatives in health care services: issues and findings. Int J Product Perform Manag 56:7–24
Lasorsa I, Liuzzi G, Calabrese R, Accardo A (2015) An innovative method for standardizing lean management approach in hospitals. In 16th Nordic-Baltic conference on biomedical engineering. Springer, pp 67–70
Mann D (2014) Creating a lean culture: tools to sustain lean conversions. CRC Press, Boca Raton
Niemeijer GC, Flikweert E, Trip A, Does RJ, Ahaus KT, Boot AF, Wendt KW (2013) The usefulness of lean six sigma to the development of a clinical pathway for hip fractures. J Eval Clin Pract 19:909–914
Niemeijer GC, Trip A, Ahaus KT, Does RJ, Wendt KW (2010) Quality in trauma care: improving the discharge procedure of patients by means of Lean Six Sigma. J Trauma Acute Care Surg 69:614–619
Ohno T (1988) Toyota production system: beyond large-scale production. Productivity Press, Portland
Oppenheim BW (2011) Lean for systems engineering with lean enablers for systems engineering. Wiley, Hoboken
Paul JA, Jotshi A (2013) Efficient operating room redesign through process improvement and optimal management of scheduled and emergent surgeries. Int J Math Oper Res 5:317–344
Pettigrew A, Ferlie E, Mckee L (1992) Shaping strategic change-the case of the NHS in the 1980s. Public Money Manag 12:27–31
Radnor Z, Walley P, Stephens A, Bucci G (2006) Evaluation of the lean approach to business management and its use in the public sector. Scottish Executive Edinburgh, Scotland
Radnor ZJ (2010) Review of business process improvement methodologies in public services. The Advanced Institute of Management Research, London
Radnor ZJ, Holweg M, Waring J (2012) Lean in healthcare: the unfilled promise? Soc Sci Med 74:364–371
Rees GH (2014) Organisational readiness and lean thinking implementation: findings from three emergency department case studies in New Zealand. Health Serv Manag Res 27:1–9
Schwarz P, Pannes KD, Nathan M, Reimer HJ, Kleespies A, Kuhn N, Rupp A, Zügel NP (2011) Lean processes for optimizing OR capacity utilization: prospective analysis before and after implementation of value stream mapping (VSM). Langenbeck's Arch Surg 396:1047–1053
Shah R, Goldstein SM, Unger BT, Henry TD (2008) Explaining anomalous high performance in a health care supply chain*. Decis Sci 39:759–789
Siddique K, Elsayed SEA, Cheema R, Mirza S, Basu S (2012) One-stop cholecystectomy clinic: an application of lean thinking-can it improve the outcomes? J Perioper Pract 22:360–365
Simon RW, Canacari EG (2014) Surgical scheduling: a lean approach to process improvement. AORN J 99:147–159
Simons FE, Aij KH, Widdershoven GA, Visse M (2014) Patient safety in the operating theatre: how A3 thinking can help reduce door movement. Int J Qual Health Care 26:mzu033
Singh J, Singh H (2009) Kaizen philosophy: a review of literature. Icfai Univ J Oper Manag 8: 51–72
Sinha KK, Kohnke EJ (2009) Health care supply chain design: toward linking the development and delivery of care globally*. Decis Sci 40:197–212

Sloan AF, Hayes KJ, Radnor Z, Robinson S, Amrik Sohal T, Al-Balushi S, Sohal A, Singh P, Al Hajri A, Al Farsi Y, Al Abri R (2014a) Readiness factors for lean implementation in healthcare settings–a literature review. J Health Organ Manag 28:135–153

Sloan AF, Hayes KJ, Radnor Z, Robinson S, Amrik Sohal T, Drotz E, Poksinska B (2014b) Lean in healthcare from employees' perspectives. J Health Organ Manag 28:177–195

Sobek II DK, Jimmerson C (2006) A3 reports: tool for organizational transformation. In IIE annual conference. Proceedings, 2006. Institute of Industrial Engineers-Publisher, 1

Teichgräber UK, De Bucourt M (2012) Applying value stream mapping techniques to eliminate non-value-added waste for the procurement of endovascular stents. Eur J Radiol 81:e47–e52

Tolga Taner M, Sezen B, Antony J (2007) An overview of six sigma applications in healthcare industry. Int J Health Care Qual Assur 20:329–340

Torabi SA, Giahi R, Sahebjamnia N (2016) An enhanced risk assessment framework for business continuity management systems. Saf Sci 89:201–218

Toussaint J (2009) Writing the new playbook for US health care: lessons from Wisconsin. Health Aff 28:1343–1350

Van Oostrum JM, Van Houdenhoven M, Hurink JL, Hans EW, Wullink G, Kazemier G (2008) A master surgical scheduling approach for cyclic scheduling in operating room departments. OR Spectr 30:355–374

Van Vliet EJ, Sermeus W, Van Gaalen CM, Sol JC, Vissers JM (2010) Efficacy and efficiency of a lean cataract pathway: a comparative study. Qual Saf Health Care 2008:028738

Villa D (2010) Automation, lean, six sigma: Synergies for improving laboratory efficiency. J Med Biochem 29:339–348

Waldhausen JH, Avansino JR, Libby A, Sawin RS (2010) Application of lean methods improves surgical clinic experience. J Pediatr Surg 45:1420–1425

Waring JJ, Bishop S (2010) Lean healthcare: rhetoric, ritual and resistance. Soc Sci Med 71:1332–1340

Wickramasinghe N, Al-Hakim L, Gonzalez C, Tan J (2013) Lean thinking for healthcare. Springer Science & Business Media, New York

Womack JP, Jones DT, Roos D (1990) Machine that changed the world. Simon and Schuster, New York

Wu S, Liu P, Belson D (2010) Application of lean principles to improve congestive heart failure care. In IIE annual conference. Proceedings, 2010. Institute of Industrial Engineers-Publisher, 1

Wysocki B (2004) To fix health care, hospitals take tips from factory floor. Wall Street J (April 9):A1

Yousri T, Khan Z, Chakrabarti D, Fernandes R, Wahab K (2011) Lean thinking: can it improve the outcome of fracture neck of femur patients in a district general hospital? Injury 42:1234–1237

Zidel TG (2006) A lean toolbox: using lean principles and techniques in healthcare. J Healthc Qual 28:W1–W7

Chapter 22
Procurement Management in Healthcare Systems

Abbas Ahmadi, Mir Saman Pishvaee, and S. Ali Torabi

22.1 Introduction

The healthcare sector is the world's largest industry in terms of budget, employees, customers, etc. For example, healthcare sector in the USA accounts for 16% of GDP (Ketikidis et al. 2010; Kirsch 2002). This industry, such as the others, requires an efficient management of its procurement affairs in order to move in a perfect way. Procurement system constitutes the mainstay of any organization and significantly influences the other core and supportive systems. This importance is growing for the healthcare systems since they are facing with extreme political and public pressures.

Procurement affairs significantly affect the overall performance of a supply chain and influence all activities performed throughout organizations. Hereupon, Procurement management is usually considered as the most valuable part in supply chain management, particularly in the healthcare industry. Monczka et al. (2008) state that the ratio of purchasing costs of organizations to their revenues is on average 55%. In addition, 50–90% of their financial flow is associated with procurement matters. Also, organizations typically pay high salaries to personnel of procurement departments, due to their crucial roles in achieving the goals (Monczka et al. 2008). Therefore, this function is an important part of management efforts to reduce costs and to enhance the competitive advantage (Ketikidis et al. 2006; Talluri et al. 2006).

A. Ahmadi • M.S. Pishvaee
School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran
e-mail: ab_ahmadi@ind.iust.ac.ir; pishvaee@iust.ac.ir

S.A. Torabi (✉)
School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran
e-mail: satorabi@ut.ac.ir

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3_22

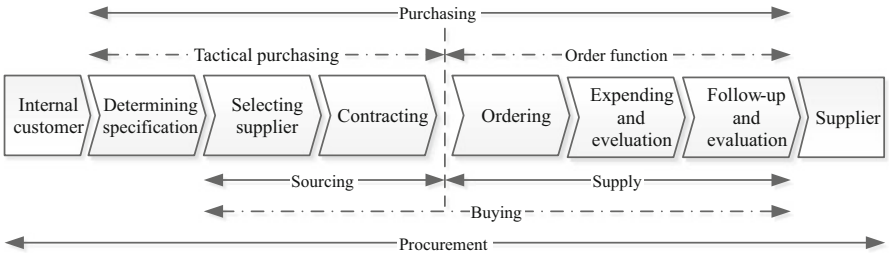

Fig. 22.1 Representative view of different definitions (Adapted from Van Weele (2009))

22.1.1 Definitions and Scope

In the literature, four terms of *purchasing*, *procurement*, *supply* and *sourcing* are employed interchangeably (Leenders et al. 2001; Van Weele 2009), while their definitions are somewhat different. However, they are similar in some respects. For instance, the phrase *obtaining external products or services* are common in all definitions (Fearon et al. 1992; Lysons and Farrington 2006; Knudsen 1999).

Van Weele (2009) reviewed various definitions of the above-mentioned words and expressed a framework in which the differences and similarities between them are highlighted (see Fig. 22.1). As Fig. 22.1 shows, procurement includes a wide variety of activities required to deliver the products from suppliers to final destinations. Procurement also covers the other relevant concepts such as purchasing, supply, and sourcing. In this definition, purchasing refers to all activities for which the organization receives an invoice from outside parties. In other words, purchasing department pursues to provide the required inputs at the right cost, in the right quantity, with the right quality, at the right time and from the right source. However, procurement drives the organization to achieve its objectives, in addition to supporting the aforementioned activities (Knudsen 1999). Procurement may arise at the local, national and international levels among a number of public, private, national and local entities (Rao et al. 2006). The US government (Government US 1996) defined the procurement as "the process of obtaining services, supplies, and equipment in conformance with applicable laws and regulations".

The procurement process is intrinsically complex, especially in healthcare systems. In fact, the healthcare is naturally different from other sectors, because the concerns of healthcare providers must be focused on managing both consumers (i.e., patients) and payers (i.e., insurance providers and government agencies) simultaneously. The complexity is due to the large number and the great variety of contributors for accomplishing the work. Indeed, the coordination of health agencies, suppliers, manufacturers, funding sources, and other entities involved requires to deploy sophisticated procedures. This issue becomes more prominent when the organization faces with limited human resources, lack of information about prices and suppliers, inadequate financing, an absence of awareness of government and donor regulations, overlapping systems, and unsynchronized or

outdated rules and guidelines. Such challenges may result in high prices, delayed shipments and eventually reduced access to essential medicines for patients (Rao et al. 2006; Klein 2012).

Procurement management involves several important decisions, which will be addressed in this chapter. These decisions can perfectly be handled through efficient, but sophisticated optimization techniques, which can lead to outstanding performance of the system. Hence, this chapter generally focuses on identifying the decision-making problems within the context of procurement management in the healthcare systems, as well as applied mathematical modeling approaches to formulate these problems and find their optimal solutions. In this regard, the next section discusses about the pressures and limitations ahead of the procurement management, besides introducing the healthcare supply chain. Section 22.3, which constitutes the main body of the chapter, is dedicated to the most popular procurement related decision problems in the healthcare systems and their corresponding optimization models. Section 22.4 provides a real-life case study to present the usefulness and practical value of optimization models in the context of healthcare procurement management. Finally, a number of attractive future research directions are presented in Sect. 22.5.

22.1.2 Importance, Advantages and Drivers

The advances of medical, biological and technological engineering sciences alongside the growth and ageing of the population have resulted in the healthcare cost inflation. Therefore, the management departments of healthcare organizations are trying to recognize the most effective and efficient ways of reducing the healthcare costs, while at the same time providing better and safer services to patients (Okoroh et al. 2001; Cunningham and Wilson 1992; Fadlalla and Wickramasinghe 2004; Mousazadeh et al. 2015). These issues are more prominent in the countries where the rate of ageing population is rapidly increasing over time, which impose the higher health care expenditure. For example, the elderly population of Singapore is estimated to increase from 11% to 27% by the year 2030 (Kumar et al. 2005). Also, according to the report of WHO (world health organization) in 2000, fully one-third of the world's population is deprived of indispensable health services. Especially, in 2001, it was revealed that about 50% of people in South Asia and Sub-Saharan Africa does not have access to the basic health services.

In the viewpoint of Kirsch (2002), the healthcare industry by having more than three trillion dollars of budget and tens of million employees can be considered as the largest industry in the world. In all the developed and developing countries, this industry is generally the most important and considerable segment of the service sector (Ketikidis et al. 2010). For example, the healthcare market constitutes the largest service industry in the USA (Smith and Correa 2005).

Ford and Scanlon (2007) stated that the US healthcare sector is faced with many quality improvements and particularly cost containment challenges. After the labor,

material cost constitutes the major source of expenditure in healthcare, such that the cost of goods and supplies are expected to be 16–28% of a hospital's budget (Duclos 1993). These facts have ensued to great political and public pressures on the healthcare industry to control the costs of medicine (Gary Jarrett 1998; Lambert et al. 1997). Henceforth, it is vital for healthcare providers to control expenditures and prices in an efficient way. This event also strongly encourages the providers to compete with each other on the prices offered to the customers (Gary Jarrett 1998).

In addition to price concerns, the following items can be mentioned: (1) waste of time and resources for seeking the items required (while on the other hand, manufacturers and distributors are desperately searching for customers); (2) confronting with inefficiency issues (especially in pharmaceutical industry); (3) declining customers confidence in the healthcare system's ability to meet fundamental delivery needs; (4) concerns over clinical quality and patient safety; (5) a growing number of uninsured and underinsured; (6) the significant and growing percentage of the gross domestic product spent on healthcare; and (7) failure to safeguard the availability of supplies and consequently to the health delivery (Ford and Scanlon 2007; Gary Jarrett 1998).

All aforementioned facts reveal the extreme need to reduce healthcare costs and obviating other challenges without scarifying the quality of services; i.e. achievement the goal of "providing healthcare efficiently and cost-effectively".

According to the viewpoint of many researchers and economists, the significant increase in healthcare costs is due to the inefficient procurement affairs (Chen et al. 2004; Carr and Smeltzer 1997; Pohl and Förstl 2011; Baier et al. 2008). The financial performance of any organization is essentially influenced by their purchasing and procurement activities. Therefore, a healthcare organization should establish an effective procurement strategy in order to achieve a super normal performance (i.e., a performance above the whole sector's average performance). Notably, procurement strategy must be consistent with other objectives and strategies of the corporate (Chen et al. 2004; Carr and Smeltzer 1997; Pohl and Förstl 2011; Baier et al. 2008).

In the absence of an effective procurement plan, the healthcare systems may face with various problems. For example, in the case of inadequate procurement capacity within the medicine supply chain, the system may experience stock-outs, low-quality medicines or supplies, or even expired or counterfeit goods. Such threats can place people at risk, undermine confidence in public health services and drain limited resources (Simonsen et al. 1999).

22.2 Characteristics of Procurement Management in Health Systems

22.2.1 Limitations: Public and Political Pressures

As previously mentioned, the government and insurers have driven the healthcare industry to hinder the rapidly increasing costs of treatment and other core activities

of health systems. In this respect, Gary Jarrett (1998) asserted that during the last three decades, the healthcare providers have been under extreme political and public pressures to control the relevant costs. The author claimed that the achievement of this aim entails imposing limitations either on prices, on quantities of services, or on both. However, in this chapter it will be indicated that these scenarios are not the only ways to reach desirable outputs in a health system. Indeed, some decision making tools and techniques can be employed to improve the health systems in multiple dimensions, simultaneously (e.g. The use of mathematical programming and game theory based approaches in the context of purchasing procedures and cooperation among different players in the procurement affairs).

However, the healthcare systems always deal with many challenges which impose high pressure to employ more effective cost control tools and seeking for finer methods. These methods should be designed so that to enable the efficient management of scarce resources, to consider health outcomes and population health gain, to become more responsive to issues and concerns of patients, to increase service quality while strengthening operating efficiencies, etc. (Saltman et al. 1998; Issel 2005). For instance, researchers and policymakers are concerned that the US healthcare sector is failing to deliver health services effectively and/or efficiently. In response, some large organizations are reassessing their approaches to purchasing health insurance and healthcare services for their employees (Ford and Scanlon 2007). It is worthy to know that the report of WHO on world health (WHO 2000) highlighted the strategic purchasing as a major means for improving performance of health systems in various aspects (Figueras et al. 2005).

22.2.2 Procurement in Healthcare Supply Chains

Papageorgiou (2009) stated that a supply chain (SC) is "a network of facilities and distribution mechanisms that handles the functions of material procurement, material transformation to intermediates and final products, and distribution of these products to customers". The concept of supply chain management (SCM) originates from manufacturing and industrial context. SCM covers three main flows of product, information and finances. Therefore, there is a need for a procurement attitude which seeks to increase the effectiveness and efficiency of product delivery by means of coordinating the relationship among suppliers and purchasers (Ford and Scanlon 2007).

To date, SCM has extensively been employed in various fields including healthcare. Nevertheless, the healthcare area still feels a great deficiency in this context; especially as it is believed that applying the SCM to healthcare systems is more complex in comparison to manufacturing systems. This matter is further highlighted in procurement stage, because the portfolio of healthcare products is substantial, the healthcare providers' requirements overlap considerably, and vendor and contract management are quite challenging (Saha et al. 2011).

Since a considerable proportion of health service finished cost is related to the supplied materials, procurement management can be considered as the most

valuable part of healthcare SCM. Therefore, procurement management is a streamlined tool for the firms to reduce costs (to control purchasing process) and also to enhance the competitive advantage (Ketikidis et al. 2006; Talluri et al. 2006). Indeed, it is easy to inspire ideas from the techniques presented to approach the SCM problems in procurement management in order to provide better management to decrease costs and to increase quality of organizational inputs, and ultimately deliver a better performance over the organizations' affairs. In this respect, since many years ago, the healthcare sector (including coalitions, employers, the federal government, etc.), such as other organizations, has utilized the SCM techniques to meet their needs in procurement of medical supplies (Burns 2002; Ford and Hughes 2007).

Recently, a number of intermediary entities have emerged in the healthcare supply chain, which facilitate the connection between appropriate buyers and sellers. In fact, the presence of complexity in the procurement process and the potential advantage of economies of scale encourage the healthcare providers to outsource parts of their purchasing activities and to cooperate with each other to utilize the relevant benefits. These intermediary players are known as group purchasing organizations (GPOs). Given the specific nature of healthcare SCs and the prevalent challenges they encounter, GPOs will play a critical role in today's healthcare SCs (Saha et al. 2011; Nollet and Beaulieu 2003;).

22.3 Healthcare Procurement Management Problems and Modelling Methods

The aim of this section is to present the problems prevalent in the healthcare system procurement management. These problems are introduced in separate subsections, according to their subjects. However, the problems can be categorized according to the applied modeling techniques. These techniques are mainly mathematical programming (including MILP[1], NLP[2], and MINLP[3]), game theory, and other quantitative methods (such as AHP[4]). In order to create a solid and comprehensive view of the matter, each decision problem is followed by the related modeling approaches used in the literature.

22.3.1 Planning Problems Related to Group Purchasing Organizations (GPOs)

Nowadays, organizations have a tendency to aggregate their purchasing volumes through purchasing groups to gain economies of scale and other relevant advantages.

[1] Mixed integer linear programming.

[2] Non-linear programming.

[3] Mixed integer non-linear programming.

[4] Analytic hierarchy process.

The terms such as collaborative purchasing, collective purchasing, pooled purchasing, joint purchasing, group purchasing, consortium purchasing, horizontal cooperative purchasing, shared purchasing, bundled purchasing and alliance purchasing are interchangeably used in the literature. This concept basically refers to the horizontal cooperation between two or more organizations in one or more steps of their purchasing process, which is accomplished through pooling and/or sharing their purchasing volumes, information, market and demand risks and/or resources (Burns and Lee 2008; Bakker et al. 2008; Schotanus et al. 2010).

The concept of group purchasing has experienced four different time periods during its evolution. Until the 1970s, purchasing groups were mostly established by local or regional hospitals and the healthcare institutions held by religious groups or by governmental agencies. Indeed, this novel concept has originated from the healthcare industry. In the next period, which began at the end of the 1970s, the national groups emerged. They were primarily private companies, which their main function was to combine the volume of purchases to offer lower prices. A large number of corporations were formed in this period. In the 1980s, due to the great growth of such companies, a competition was formed among them. This new situation motivated the groups to expand their services, besides contract negotiation, to consulting, computer services, vehicle rentals, and so on. Ultimately, at the fourth period (i.e. in the early 1990s), small purchasing groups were consolidated and subsequently mega-groups emerged (Nollet and Beaulieu 2003).

A GPO is a third party organization which takes over the matters relating to purchases of the healthcare sector. Interested readers can refer to (Adobor and McMullen 2014), to gain a comprehensive insight about the third party entities in strategic procurement and supply chain management. According to the definition of health care supply chain association (HSCA 2011a) "a GPO is an entity that helps health care providers (such as hospitals, ambulatory care facilities, nursing homes and home health agencies) realize savings and efficiencies by aggregating purchasing volumes and using that leverage to negotiate discounts with manufacturers, distributors and other vendors". The costs incurred by healthcare providers are carried out by their customers (i.e., patients). The healthcare providers often waste their time and resources in searching for the required items meeting their criteria (Ketikidis et al. 2010). While, GPOs allow healthcare providers to concentrate on their primary activities, which leads to the enhancement of the offered services' quality and cost saving (Schneller 2009). They also can deliver the services including clinical review processes, benchmarking, e-commerce solutions, and technology assessment to their members (Burns 2014). Following to realizing the essential impact of GPOs on the performance of healthcare systems, nearly 96–98% of hospitals in the US use GPOs (on average, two to four GPOs per facility) to make about 72% of their purchases (HSCA 2011a). Also, GPOs save the US healthcare industry $38 billion dollars annually (Schneller 2009).

Healthcare providers (e.g. hospitals) can become a member of a GPO through paying a membership fee specified by the GPO. The GPO proceeds to search for suppliers, evaluate them, negotiate and finally contract with the appropriate suppliers. Then, the GPO announces the members about "on contract prices" (i.e.

the prices that the GPO attained from the suppliers through negotiation). From this moment on, when a GPO member needs a certain item, it can purchase the needed item at the negotiated price from the relevant suppliers contracted with the GPO. It should be noted that the healthcare providers may belong to more than one GPO (Burns and Lee 2008), nevertheless they are free to decide whether to purchase their required items through GPO or directly from suppliers; but, they usually find the first alternative to be better deal (Saha et al. 2011). Figure 22.2 illustrates the chronological sequence of the events, which occur between GPO and healthcare providers. On the other side, the GPO receives a percentage (commonly less than 3%) of the total monetary value of the transactions carried out by the members as contract administrative fee (CAF). The GPO may also charge administrative fees to authorized distributors who deliver products to healthcare providers (HSCA 2011b). The working process of GPOs is absolutely transparent to their members, and they make all information on administrative fees available to them (Figueras et al. 2005). Figure 22.3 provides a schematic view about the aforementioned descriptions presented for the mechanism of GPO, using depiction the different relationships among the suppliers, healthcare providers, and GPOs.

Over the time, GPOs have been faced with several challenges from critics. The critics mainly claim that: (1) GPOs do not offer lower prices to their members (otherwise they gain less administrative fee from suppliers, i.e., fewer CAF), (2) CAF provokes suppliers to inflate their prices (with the aim of compensating the commission paid to the GPOs), and (3) GPO stifles innovation and competition among suppliers (because of employing a monopoly supply policy). But, as Burns (2014) asserts most of the allegations that critics have upon the GPOs performance do not have scientific foundations; even in advanced fashions, these conclusions have been inferred from the interviews and surveys conducted with hospital purchasing managers. Recently, academics and researchers (Burns and Lee 2008; Serb 2010; Hu et al. 2012; Hu and Schwarz 2011) have strived to examine these challenges in the view of analytical approaches. In summary, the findings and relevant evidences indicate that group purchasing is a strategic tool used worldwide by different types of hospitals in both public and private sectors. There are many advantages associated with the presence of GPOs. For instance, they reduce both hospital costs and national spending, increase hospital awareness of new products and technologies, foster competition between suppliers, and so on (Burns 2014).

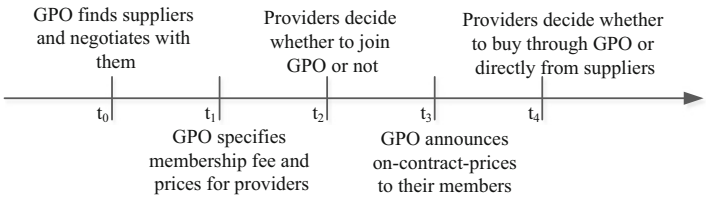

Fig. 22.2 The sequence of events between GPO and healthcare providers (Inspired from Saha et al. (2011))

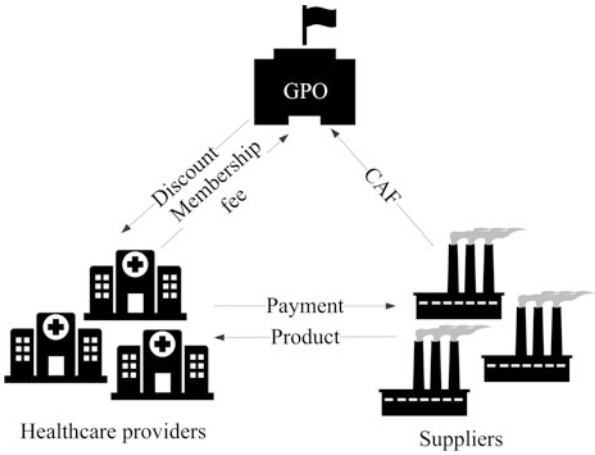

Fig. 22.3 A schematic view of the interactions among different players in a GPO-based purchasing system (Inspired from Saha et al. (2010))

However, they might have some drawbacks, of course not those of the former critics have mentioned. Rego et al. (2014) exhaustively summarizes the advantages and disadvantages of the presence of GPOs in healthcare SCs (see Fig. 22.4).

Few academic studies, particularly with analytical approaches, have been conducted about GPOs to date. Among these, Nagarajan et al. (2009) deployed the cooperative game theory to design a fair framework for allocating the benefits gained through the coalition of the members, based on their contribution in marginal value. This framework is able to enhance the stability of the alliance as well as eliminating of some limitations of the rules commonly used in practice for sharing the savings. In another research, Marvel and Yang (2008) considered nonlinear tariffs (instead of linear ones that generally used in the literature), in the case that suppliers have the high bargaining power, in order to intensify the competitive effects of a non-for-profit GPO and eliminate the allocative inefficiency of single price competition. Recently, Saha et al. (2011) studied the impact of market price uncertainty on the decisions taken by the healthcare providers (buyers) and the suppliers (sellers). The problem was investigated under two different scenarios. In the first scenario which was named "with custom pricing", suppliers could offer lower prices than the GPO-negotiated price to healthcare providers. However, in the second scenario, which was called "without custom pricing", this option is prohibited. Indeed, some of the healthcare providers (especially the large ones) may join to GPO only for the purpose of retrieving the GPO price as a starting point for further negotiation and also for reducing the price uncertainty. This opportunity is prepared in the first scenario. Blair and Durrance (2013) also presented an economic analysis around the competitive impact of GPOs and their contracting and funding mechanisms.

To the best of our knowledge, Burns (2002) was the first one who considered the concept of the GPO in the healthcare supply chain. He analyzed the role of GPOs in the supply chain of the US healthcare industry, as well as their operations and strategies. A few years later, the second study was performed by Schneller and Smeltzer (2006). It is notable that both of the performed studies are based

	Advantages	Disadvantages
For healthcare providers	• Reduction of cost related to purchasing, transaction, and administration • Human resources savings (because of outsourcing some purchasing efforts) • Increased information on supply markets • Increased focus on core operational activities	• Decreased ability to fulfil the needs of decentralized users (due to standardization) • Lower innovation capabilities at contract and product/service levels (due to compromise, standardization and reduction of direct contacts with suppliers) • Lower responsiveness in emergency situations (e.g. for a case in small scale)
For GPOs	• Increased negotiation and bargaining power versus suppliers (due to consolidation of purchase volumes) • Lower unit cost (due to increased economies of scale) • Reduction of the number of transactions and so reduction of duplicated purchasing efforts • Development of purchasing expertise • Rationalized choice through better-informed selection and standardization • Improved ability to respond to large scale emergency situations (due to increased flexibility of inventories, coordination, and resource pooling)	• Imposing coordination costs, mainly when GPO size increases

Fig. 22.4 Advantages and disadvantages of GPOs (Adapted from Rego et al. (2014))

on qualitative methods and dealt with the performance, challenges and criticisms of the GPOs in the healthcare supply chains. Recently, Hu et al. (2012) and Hu and Schwarz (2011) accomplished analytical investigation on healthcare supply chains through the game theory and Hotelling model, respectively. The supply chain considered by Hu and Schwarz (2011) involved a continuum of identical healthcare providers (with the same purchasing requirement), a non-for-profit GPO (which formed by healthcare providers), and two manufacturers (offering competing but not identical products). In contrast, the supply chain studied by Hu et al. (2012) has a discrete number of healthcare providers which might have different requirements; one manufacturer and also a competitive source who was selling the product with a fixed price. Here, the studied GPO was assumed to be a for-profit entity.

Last but not the least, Rego et al. (2014) examined the problem of designing and also evaluating the structure of GPOs for a set of healthcare providers who cooperate with each other in a supply chain. They attempted to determine the number of GPOs to form, their sizes and composition. They also established the material flow through

the supply chain (i.e. where, when and in which quantities the supplied items are stored and flow in the supply chain). They developed a hybrid optimization approach involving VNS and Tabu-Search metaheuristic algorithms to solve the extended MINLP model.

22.3.1.1 A Game Theoretic-Based GPO Problem

Hu et al. (2012) proposed a game theoretic-based model with complete information (i.e. all players are informed about the payoff of each other) to deal with the problem in which a GPO is engaged. The model involves $n + 2$ players, including a profit-maximizing GPO, a profit-maximizing manufacturer (with which the GPO has contracted), and n healthcare providers. Also, it is assumed that there is an external competitive source who sells the product at a fixed exogenous price.

The healthcare providers' demands (i.e. their purchasing requirements) for each product are different to each other, in terms of type and size, and they are known for each one. When the healthcare providers join the GPO, they are free to use the GPO contract or negotiate directly with the suppliers. Indeed, each healthcare provider can split its requirement among the GPO, the manufacturer, and the competitive source in order to minimize its total purchasing cost. Here, in order to model the problem, it is assumed that the GPO buys the products from the manufacturer at a price and sells them to its members with another price. Nevertheless, in reality, GPOs act as an intermediary; they neither buy nor sell products, as was already mentioned.

The manufacturer offers a same quantity discount schedule for both the GPO and healthcare providers. Accordingly, the GPO determines the prices for their members. In consequence, each healthcare provider decides on the amount that supplies from each party (i.e. the GPO, the manufacturer, and the competitive source). The following notations are used to mathematically formulate the problem:

Parameters

q_i	Fixed purchasing requirement of healthcare provider i ($i = 1, 2, \ldots, n$)
$\widehat{p}$	The fixed unit cost of the competitive source
f^M	The fixed contracting cost for each healthcare provider, when purchasing from the manufacturer or competitive source
$\widehat{f}^G$	GPO membership fee
$\tilde{f}^G$	The fixed contracting cost for each healthcare provider, when purchasing through the GPO
f^G	$= \widehat{f}^G + \tilde{f}^G$, which it is assumed ($f^G \leq f^M$)
λ	CAF ($0 \leq \lambda \leq 1$)

Variables

u_i	1 if healthcare provider i purchases through the GPO and 0 otherwise
v_i	1 if healthcare provider i purchases from the manufacturer and 0 otherwise
w_i	1 if healthcare provider i purchases from the competitive source and 0 otherwise
x_i	Purchase quantity of healthcare provider i through the GPO
y_i	Purchase quantity of healthcare provider i from the manufacturer
z_i	Purchase quantity of healthcare provider i from the competitive source
p^G	GPO's per unit on-contract price
$p(\cdot)$	Manufacturer's quantity-discount schedule: for a quantity q, the manufacturer offers a price of $p(q)$

For simplicity, it is assumed that the healthcare providers are indexed according to their non-decreasing purchasing requirements, i.e. $q_1 \leq q_2 \leq \ldots \leq q_n$. Now, the aforementioned game can conceptually be described as the following optimization problem for each player. *For healthcare provider* $i (i = 1, 2, \ldots, n)$:

$$\pi_i = \min \left(f^G u_i + p^G x_i\right) + \left(f^M v_i + p\left(y_i\right) y_i\right) + \left(f^M w_i + \widehat{p} z_i\right) \tag{22.1}$$

s.t.

$$x_i + y_i + z_i = q_i \tag{22.2}$$

$$x_i \leq q_i u_i \tag{22.3}$$

$$y_i \leq q_i v_i \tag{22.4}$$

$$z_i \leq q_i w_i \tag{22.5}$$

$$u_i, v_i, w_i \in \{0, 1\}; x_i, y_i \geq 0 \tag{22.6}$$

Objective function (22.1) shows the cost of purchasing from the GPO, manufacturer and competitive source, respectively. Constraint (22.2) is the demand satisfaction of the healthcare provider i. Constraints (22.3–22.5) indicate that the healthcare provider i can purchase its requirements from each party, when a contract is established between them. Constraint (22.6) shows the type of decision variables. Also, the problem can be formulated from the viewpoint of the GPO as follows: *For the GPO*:

$$\pi_G = \max \widehat{f}^G \sum_{i=1}^{n} u_i + p^G \sum_{i=1}^{n} x_i - (1 - \lambda) p \left(\sum_{j=1}^{n} x_j\right) \sum_{i=1}^{n} x_i \tag{22.7}$$

s.t.

$$p^G \geq 0 \tag{22.8}$$

Objective function (22.8) is composed of revenues from membership fees and on-contract sales, as well as the costs, respectively. Constraint (22.8) limits the GPO price to being a positive value. Finally, for the manufacturer, we have: *For the Manufacturer*:

$$\pi_M = \max (1 - \lambda) p \left(\sum_{j=1}^{n} x_j \right) \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} p (y_i) y_i \tag{22.9}$$

s.t.

$$\begin{array}{l} p(q) \text{is non} - \text{increasing in } q \\ p(q)q \text{ is non} - \text{decreasing in } q \end{array} \tag{22.10}$$

Objective function (22.9) denotes that the manufacturer gains revenue through the GPO (discounted by the CAF) or directly from the healthcare providers. In Constraint (22.10), it is assumed that the presented quantity discount and the total purchasing cost are respectively non-increasing and non-decreasing in quantity (following the assumption $q_1 \leq q_2 \leq \ldots \leq q_n$ which taken previously). Notably, $p(q)$ is the unit price (based on the quantity discount for q unite of products). Consequently, $p(q) \,.\, q$ represents all purchasing cost for q unit of products.

Hu et al. (2012) described the optimal strategies of the problem through three lemmas and then provided a procedure to obtain the sub game-perfect Nash equilibrium (SPNE). The readers are referred to the mentioned reference to find more details about this procedure as well as the lemmas. Generally, in the SPNE, each healthcare provider purchases all the requirement form only one party (i.e. the GPO, the manufacturer or the competitive source) and does not split it among different sources. Also, it is optimal for the GPO to set its unit on-contract price to a breakeven price. The optimal strategies for the healthcare providers and the GPO depend on the quantity discount provided by the manufacturer. Moreover, it is proved that the optimal decision of a healthcare provider about purchasing through the GPO (manufacturer), in equilibrium, is also optimal for all healthcare providers with smaller (larger) purchasing requirements.

22.3.2 Supplier Selection and Order Splitting

Supplier selection is generally dealing with determining suitable suppliers having acceptable features to supply the required products and services. This area is quite rich and has extensively been studied in the SCM literature. Many techniques have

been introduced to select (and rank) the right supplier(s) according to a number of criteria. After ranking the suppliers, two situations arise: (1) working with a single supplier (i.e. following single-sourcing policy) or (2) contracting with multiple suppliers (i.e. following multiple-sourcing policy). In the first case, all orders are allocated to only one supplier, but in the second one, another decision must be made about the order quantity which should be allocated to each selected supplier. The latter decision is often called order splitting (or order allocation), which is discussed in more detail later. In summary, four major decisions are often required to be taken in multiple-sourcing case: What products or services to order, from which suppliers, in what quantities, and in which time periods (Songhori et al. 2011)?

The procedure of selecting appropriate suppliers is composed of four main steps: (1) defining the most important objectives of the supplier selection, (2) identifying the appropriate selection criteria, (3) preselecting qualified suppliers, and (4) ranking the qualified suppliers (De Boer 1998).

There may be many reasons and objectives beyond the supplier selection. The policy adopted by the company may require the buyers and their internal customers to identify a minimum number of appropriate potential suppliers for every product or service. Also, the presence of several suppliers can result in driving competition among suppliers, reduced supply risks, and attainment of many other business objectives such as supplier diversity (Beil 2009). Generally, this issue thoroughly depends on the considered industry, the product types, and the policies and strategies adopted for the organization.

After determining the objectives, the key evaluation criteria should be defined. Various criteria may be employed to assess the suppliers in terms of purchasing and procurement performance, however, several ones such as quality, price, delivery, and service consistently emerge in different situations (Lambert et al. 1997).

According to the study conducted by Swift (1995) on different industries and product types, it is suggested that in multiple sourcing situations, the key parameters are quality, price and delivery, whereas technical support and reliability of the product/service are more important in the case of single sourcing. These concepts have not been well developed in the health industry. This fact is due to the special criteria prevalent in the health sector. Several healthcare-related criteria have been identified by Lambert et al. (1997), Lehmann and O'shaughnessy (1974), and Stentoft Arlbjørn and Pazirandeh (2011). Lambert et al. (1997) studied various criteria for supplier selection in the healthcare industry and ranked them according to their importance. The research team concluded that in spite of the remarkable pressures of government and insurers on healthcare providers to reduce costs, low price does not give higher priority than product quality, delivery, and service in terms of importance. Moreover, they found that, in spite of defining various criteria for supplier selection, on average, the suppliers are not meeting customer requirements for key criteria. Also, Lehmann and O'shaughnessy (1974) realized that the product category affects the relative importance of supplier selection criteria.

For pre-selection of potential suppliers (i.e., the third step in the aforementioned procedure), a number of techniques such as Data Envelopment Analysis (DEA), Cluster Analysis (CA), Analytic Hierarchy Process (AHP), and Case-Based Rea-

soning (CBR) are usually deployed. Also, in the fourth stage, mathematical and statistical models, Total Cost of Ownership (TCO) method, linear weighting models, and Artificial Intelligence (AI) based models are commonly used (De Boer 1998). In order to find a comprehensive review on supplier selection, the readers are referred to (De Boer et al. 2001; Ho et al. 2010; Aissaoui et al. 2007).

The models prevalent in the literature of this area are commonly based on mathematical programming formulation (e.g., mixed integer programming) with the objective functions such as minimizing total purchasing costs and the constraints for satisfying lead time, quality and service requirements. In this respect, Gurgur (2013) investigated the supplier selection alongside the order allocation for a hospital demanding a variety of healthcare products and equipment to deliver health services to its customers. Two sets of suppliers were available for this purpose which offered either new or used (refurbished) products. The new products are more expensive than the refurbished ones, but can be supplied at any time and in any quantity. Due to safety concerns, the refurbished products must receive certification from regulator to be dispersed; hence they are limited in terms of quantity and supply time. Suppliers make available the products in the bundle mode, i.e., in the packages containing a variety of items which some of them may not be required by the hospitals.[5] Gurgur (2013) presented an MILP model with the objective function containing the costs associated with purchasing refurbished and new products and their overstocking, as well as establishment of relationships with suppliers. This model was then extended to the case in which the demand is faced with stochastic uncertainty. For this purpose, a chance constraint programing (CCP) approach was deployed. Finally, the problem was solved using a simulated annealing (SA) algorithm.

Beşkese and Evecen (2012) prepared an Analytic Hierarchy Process (AHP) framework to provide a ready-to-use model to select the most appropriate suppliers for healthcare institutions. They implemented this framework on a real-life case. Also, Benyoucef and Canbolat (2007) proposed a fuzzy AHP-based framework in the e-procurement environment for supplier selection of a hospital. The purpose of this approach was to capture the uncertainty in the decision maker's mind, when making comparisons between criteria or alternatives.

With the emergence of just-in-time (JIT) philosophy, the manufacturers tend to use single-sourcing policy, owing to the quality concerns. However, it is still popular to use multiple sources because of its numerous benefits such as reducing uncertainty, stimulating competition among suppliers, achievement of sufficient capacity, etc. (Qi 2007; Minner 2003). In this respect, an important decision to be made is how to split an order between different available suppliers. This issue is a matter of fact, especially in the procurement of medical supplies for healthcare providers which directly affect the public health. Notably, disruption risks may impose irreparable damages to the procurement process of healthcare providers' requirements (Torabi et al. 2015). For example, as an industrial case, a fire in 2000 shut down the production of the manufacturer who was supplying an electronic chip

[5]The concept of bundling products will be discussed in the next subsection.

for Ericson and Nokia. Due to the multiple sourcing strategy adopted by Nokia, it could quickly recover from the disruption; but the production of Ericson was extremely influenced due to the single sourcing (Qi 2007).

Dalkilic et al. (2006) investigated a real-case problem by considering a hospital requiring a wide variety of items, one first-tier supplier (i.e., retailer), and many second-tier suppliers procuring various, but not necessarily identical products. The hospital calls for a bidding to different retailers and the bidders are evaluated based on their prices and the offered lead times. A retailer (bidder) should decide about the prices and the lead times to be quoted in the bid. For this purpose, the retailer needs to firstly determine the materials which should be procured and secondly the appropriate supplier for each of them. Thirdly, retailer decides on the quantity and the time by which the materials should be procured (i.e., deals with an order splitting problem), in order to minimize the total cost. In this problem, it is assumed that the lead-times are stochastic. A non-linear integer programming model is developed and an improved evolutionary algorithm is deployed to solve the presented model.

22.3.3 Bundling Products

Nowadays, the healthcare providers tend to purchase refurbished durable products, instead of new ones, with the aim of achieving savings in their purchasing costs. Although, in the past, the safety reasons hindered them to use the refurbished items, now in the US, the emergence of quality and price certification issued by American Hospital Association (AHA) has eliminated such concerns. In this regard, brokers (sellers) of refurbished items are seeking to bundle their products into a variety of categories for offering to the customers (e.g., GPOs, healthcare providers, etc.), instead of selling unit products (Ross and Jayaraman 2009; Gurgur 2013). This fact is due to the great uncertainty on the demand of the available refurbished products. Indeed, the sellers may encounter with a situation in which some products are sold, but the customers do not tend to buy the rest. Therefore, bundling the products is the common strategy adopted to obviate these troubles. In the face of such a policy (i.e., offering products in bundles by brokers), the GPOs can be very helpful for healthcare providers; since the GPOs acquire the products in bundles and split them among the healthcare providers according to their requirements (Ross and Jayaraman 2009). Hence, the concept of bundling new products with refurbished products has recently received attention in the healthcare supply chain context (DeJohn 1998).

Sellers (suppliers) are confronted with the decision as how to configure the attractive and efficient bundles, i.e., what types of items and in what quantities should be incorporated into the bundle? In this respect, if the bundles are too large in quantity or high in price, smaller businesses may not bid and in contrast, if the

bundles are too small, large companies may not be interested (Ross and Jayaraman 2009). It is noteworthy to know that the concept of bundling products has been investigated previously by researchers in different areas (e.g. (Murthy et al. 2004; Van Buer et al. 1997; Rosenthal et al. 1995; McCormick et al. 2006)); however, little attention has been dedicated in the healthcare context.

Ross and Jayaraman (2009) studied supply chain which includes several suppliers who offer nonhomogeneous bundles of refurbished products, and a buying organization who decides to optimize the number of bundles purchased from the alternative suppliers. Indeed, they considered the supplier selection problem alongside the bundling problem. In addition, they considered a GPO in the problem, such that some buying organizations may have to purchase a minimum volume of items through GPO contract. This assumption imposes a restriction to the buyers to buy all items independently. An MILP model with the objective of minimizing the total cost is developed to cope with this problem.

22.3.3.1 Supplier Selection and Order Allocation of Bundled Products

In this subsection we present a model related to supplier selection and order allocation of new and bundled refurbished products. The original model is proposed by Gurgur (2013), but, here the model is represented with some little modifications.

Consider a hospital which requires procuring a variety of items (e.g. equipment). The hospital has two alternatives for the procurement of each item, i.e., (1) new and (2) used products. There are a number of suppliers who offer these items in the forms of new products, used products, or both forms. Suppliers sell the used products only within a number of bundles that may also contain the items which are not needed by the hospital. However, the refurbished products are cheaper than the new ones. The model aims to minimize the total acquisition cost of purchasing for which the following notations are used to formulate the problem.

Indices and Sets

i	Index of product types, $i \in \{1, 2, \ldots, n\}$
j	Index of suppliers, $j \in \{1, 2, \ldots, m\}$
k	Index of bundles of refurbished items, $k \in \{1, 2, \ldots l\}$
B_j	Set of bundles of refurbished items offered by supplier j, $B_j \subseteq \{1, 2, \ldots, l\}$

Parameters

c_{kj}	Unit purchase cost of bundle $k(k \in B_j)$ procured from supplier j
p_{ij}	Unit purchase cost of new product i procured from supplier j
D_i	Hospital's demand for item i
cc_i	Unit cost of overstocking for item i
δ_i	Maximum proportion of refurbished products that is allowed to be procured from all suppliers for item i
Y_{ikj}	Volume of refurbished item i in bundle $k(k \in B_j)$ from supplier j
Cap_{kj} Capacity of supplier j to procure bundle $k(k \in B_j)$	

Variables

Q_{kj}	Number of bundles $k(k \in B_j)$ procured from supplier j
X_{ij}	Number of new items i procured from supplier j
Z_j	1 if the hospital decides to source from supplier j and 0 otherwise

The intended model, which is an MILP model, can be formulated as follows:

$$\min \sum_{j=1}^{m} \sum_{k \in B_j} c_{kj} Q_{kj} + \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} X_{ij} + \sum_{i=1}^{n} cc_i \left(\sum_{k \in B_j} \sum_{j=1}^{m} Y_{ikj} Q_{kj} + \sum_{j=1}^{m} X_{ij} - D_i \right) \tag{22.11}$$

s.t.

$$\sum_{k \in B_j} \sum_{j=1}^{m} Y_{ikj} Q_{kj} + \sum_{j=1}^{m} X_{ij} \geq D_i \forall i \tag{22.12}$$

$$\sum_{k \in B_j} \sum_{j=1}^{m} Y_{ikj} Q_{kj} \leq \delta_i D_i \forall i \tag{22.13}$$

$$Q_{kj} \leq Z_j Cap_{kj} \forall j, k \tag{22.14}$$

$$Z_j \in \{0, 1\}\, Z_{ij} \geq 0\; Q_{kj} \geq 0 \forall i, j, k \tag{22.15}$$

Objective function (22.11) is reflecting the purchasing cost of the refurbished and new products, and the total cost of overstocking, respectively. In spite of the deterministic demand for each item, overstocking may still occur due to the presence

of the not required the items in the bundles or when the amount of the items in the bundles exceeds the demand. Constraint (22.12) ensures that the demand is fully met. Constraint (22.13) limits the amount of permissible refurbished items to be supplied. The maximum available capacity for each supplier is imposed by Constraint (22.14). Finally, Constraint (22.15) shows types of decision variables.

22.3.4 E-Procurement

Benyoucef and Canbolat (2007) defined electronic procurement (e-procurement) as: "any software, hardware or a combination of both, designed to facilitate the acquisition of goods by a governmental or commercial organization over the internet." The widespread use of the internet as a communication standard has motivated the development of e-procurement systems as a fundamental tool of procurement management. Hereunto, the internet has rapidly been established as a standard procurement tool in all organizations and also for many medical practices. The prominence of online procurement is obvious for everyone nowadays. We can obtain information about any given products in a short amount of time and purchase it through internet. Also, it is very easy for the sellers to directly contact with end-customers and discover the required information about them, as well as achieving a better coordination with suppliers, a higher flexibility, quicker transaction times, better supplier integration, and ultimately lower costs (Talluri et al. 2006; Essig and Arnold 2001; Presutti 2003).

E-procurement provides the opportunity for the firms to enhance the efficiency and effectiveness of purchasing activities, including sourcing, purchase order releasing, payment processing, shipment tracking, inventory control, new product specification, and information gathering, which altogether lead to reducing operational costs, shortening order fulfilment cycle time, creating collaborative partnerships, lower inventory levels, etc. The majority of the models prevalent around this theme in the literature have mainly concentrated on the justification and selection of appropriate technologies (see (Kameshwaran and Narahari 2003; Ammenwerth et al. 2003; Littlejohns et al. 2003)). Little attentions have been dedicated to addressing the issues associated with suppliers (Talluri et al. 2006).

Ketikidis et al. (2010) assert that pharmaceutical and hospital suppliers are the main target users for an efficient e-procurement procedure. This is due to the fact that health-related systems have a lot of various functions and need to be supported by the appropriate supplies. A number of the inefficiencies occurred in resources and time management in the healthcare are behind the lack of perfect information flow, which imposes extremely high costs (Ketikidis et al. 2010). Hence, the employment of information technology (IT) can easily provide an excellent opportunity to achieve an integrated and coordinated supply chain. There are many reasons for healthcare providers to utilize IT in their procurement affairs. For example, they would be aware about inventories and prices of products at any point of time, they can track their orders online, etc. Nevertheless, the healthcare SCs still have

extensively not adopted the internet and e-commerce to the same extent as other industries such as banking and retailing (Baltacioglu et al. 2007).

Silva et al. (2006) stated that the real-life procurement problems are extraordinarily complicated and consequently the exact approaches, even heuristic methods, are not able to tackle with these problems in an efficient manner. Hence, they suggested the employment of information exchange during the optimization processes. This fact somehow explains the relationship between e-procurement and optimization, as well as the reason why e-procurement related literature have been reviewed within this chapter. However, the important matter which should be noted is that this area has received little attention from the operational research community and there is good potential to be explored hereafter.

Due to the specific requirements of the healthcare industry, Ketikidis et al. (2010) proposed a prototype e-procurement optimized system (EPOS) for the healthcare sector in order to offer the advantages such as convenience, efficiency, information on new products, favorable pricing, broad selection and so on. EPOS addresses the healthcare supply chain in a unified approach and provides an innovative procurement environment through employing appropriate optimization algorithms. The solution resulted from EPOS is a reverse auction optimized procedure, in spite of the current leading e-procurement solutions which focus on the dynamic binding of buyers' requests and sellers' offers.

Nowadays, the companies are pursuing to integrate the suppliers into an e-procurement environment. However, they commonly take up the ad hoc approaches to gain this matter, which are not based on scientific grounds. As a result, their problems fall into suboptimal solutions and they may even experience a worse situation. For this reason, Talluri et al. (2006) proposed a multi period integer programming (MPIP) model by which the optimal sequence for the integration of suppliers (supply base) can be determined. This model included the elements such as implementation costs, budget limitation, purchase amounts, purchase order frequencies and integration priority. Two different scenarios were investigated in this research: (1) the firm has all e-procurement technology levels in place before integrating with its suppliers, (2) the firm continues to attain e-procurement technologies during the integration process.

22.3.5 Performance Measurement

Performance measurement is defined as the process of collecting, analyzing and/or reporting information related to the performance of an individual, group, organization, system or component. Such an assessment is conducted to see whether the outputs are in line with what was intended or should have been achieved (Wikipedia 2015; Upadhaya et al. 2014). The health industry requires performance measurement, especially in procurement processes, in order to assess their systems and evaluate the changes which can lead to improvements (Kumar et al. 2005). Also, the performance measurement can be interpreted as the third phase of the supplier

selection (as were introduced previously) and its output may be used in the fourth phase (i.e., for selecting appropriate suppliers).

In order to be able to make improvements in any system, it is necessary to have a measurement method to evaluate the system performance. The adage, "*If you don't measure it, you can't control it* "confirms this fact (Pooler and Pooler 1997). There are many methods for this purpose, which should be chosen according to the considered business situation. Measurement of procurement performance draws the ways to the firms to move in the line of continuous improvement and as a result gaining competitive advantage (Kumar et al. 2005). Within any organization, different individuals and departments may have different outlooks about what constitutes good job performance. For example, when financial controllers think of purchasing, they relate to the finance areas that affect them. Engineer scare about those technical aspects related to design responsibility, etc. (Pooler and Pooler 1997). Hence, there are many viewpoints about a single concept, and therefore, it is difficult, if not impossible, to evaluate all departments engaged with the procurement activities on the same basis. However, the lack of appropriate method for this purpose undermines the overall procurement performance, which results in higher costs for patients as well as the customer dissatisfaction in a healthcare system.

According to a survey conducted in a large number of leading manufacturing and service companies/firms, the following main measures are introduced for procurement/purchasing performance: (1) Quality of items purchased, (2) key supplier problems that could affect supply flows, (3) supplier delivery performance, (4) internal customer satisfaction, and (5) purchasing dollars devoted to inventory (Pooler and Pooler 1997). In procurement activities, quantifiable measurements such as cost savings can be evaluated easily; but as we know, whatever we deal with cannot be measured in a routine manner (e.g., customer satisfaction) (Pooler and Pooler 1997). Hence, a challenge may arise in this regard about how to quantify such qualitative characteristics? As an example for measurement of procurement performance in the healthcare, we consider and take a look at the work of Kumar et al. (2005). They developed a model by considering three components in the supply chain, collectively with the resources, procedures and outputs. Then they proposed a balanced scorecard by establishing six perspectives (i.e., customer, supplier, process, IT system, learning and growth, and overall) and a number of generic measures were mathematically formulated for each perspective. Some generic measures are as follows:

$$\text{Supplier evaluation} = \frac{\text{number of supplier evaluations that meet objectives}}{\text{the number of supplier evaluations}} \tag{22.16}$$

$$\text{Cost per order of suppliers} = \frac{\text{total expenditure of the department}}{\text{total number of purchase orders}} \tag{22.17}$$

$$\text{GPO participation rate} = \frac{\text{number of purchased items under GPO}}{\text{total number of purchased items}} \tag{22.18}$$

$$\text{Supply chain costs} = \frac{\text{total expenditure of department}}{\text{total purchase value}} \tag{22.19}$$

Kumar et al. (2005) conducted a case study and implemented their models in the purchasing department of the Singapore General Hospital, which was composed of two main sections: (1) purchasing, and (2) inventory management. The role of purchasing section was to manage all sourcing affairs, supplier selection, contracting with suppliers, ordering, and coordinating all orders to gain economies of scale. While the inventory management section was responsible for the activities related to storage and distribution of the procured items.

22.4 Case Study

In this section, a practical case study directed by Kalantari (2015) is reviewed. The authors addressed the supply chain master planning in a pharmaceutical company called ATRA, which is located in Tehran, capital of Iran. ATRA pharmaceutical company, with nearly 50 years of experience in the drug industry, is one of the subsidiary of the largest pharmaceutical investment holding, known as TPICO, in Iran. TPICO holding covers nearly half of the country's pharmaceutical demand. ATRA, is mainly engaged in producing of active pharmaceutical ingredients (APIs), food supplements, and finished drug products.

The SC of ATRA is composed of three layers of suppliers, ATRA Company, and customers (see Fig. 22.5). By concentrating on the most important elements (including products, materials, etc.), nine types of materials can be supplied from five potential suppliers scattered around the world. Indeed, the intended SC is a global SC. The materials which are supplied from the suppliers are then converted to six types of products. The products would be delivered to four main customers located in different provinces around the country (See Fig. 22.6).

According to the strategy of the company, the mid-term planning horizon involves two monthly periods. The company intends to use only two suppliers in each period, based on their performances. However, they already did not have a consistent method to make such decisions. Hence, we aim to provide an efficient framework to assist the managers to make the best decisions.

In order to achieve the above mentioned goal and propose an efficient master planning to the SC, a three stage modeling framework is presented. In the first stage, in order to evaluate the candidate suppliers, first a number of relevant criteria are identified and then a DEA (data envelopment analysis) approach is employed to rank the potential suppliers. Then, the output of the first stage is fed to a multi-objective MILP model in the second stage, which is an integrated procurement,

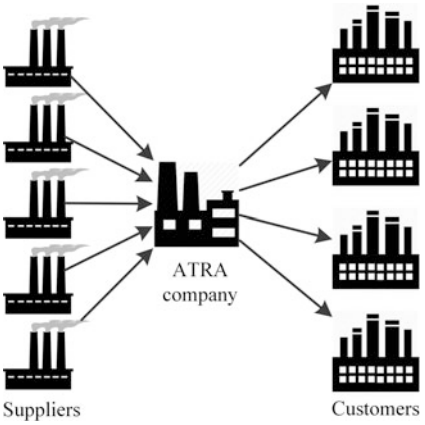

Fig. 22.5 A schematic view of ATRA's supply chain

Fig. 22.6 The situation of different players in the supply chain

production and distribution planning model over a mid-term planning horizon. The developed multi-objective MILP model accounts for epistemic uncertainty of input data for which possibilistic and robust possibilistic programming approaches are exploited. Finally, in the third stage, in order to validate the obtained master plan and also to accomplish a performance analysis of the proposed robust possibilistic programming model, a simulation process is applied. Figure 22.7 illustrates an outline of the presented framework. Since the first stage of this framework is more relevant to the subject of this chapter, it will be presented and discussed in the following. Due to space limitation, we do not elaborate on the details of the second and third stages of the proposed methodology. Interested readers may consult with Kalantari (2015) for more details.

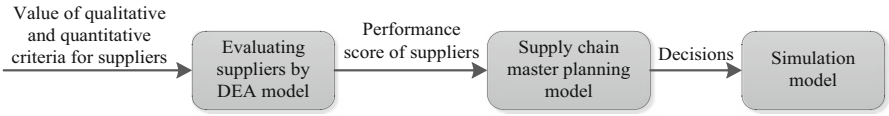

Fig. 22.7 A schematic view of the framework presented by Kalantari (2015)

Table 22.1 Data acquired for suppliers' performances in different criteria

Supplier	Purchasing price (*10^2)	Quality	Response time	Reliability in delivery	Expiration date
1	(1254,1264,1274)	(0.8,0.85,0.95)	(0.7,0.8,0.9)	(0.5,0.6,0.7)	(0.9,0.9,0.9)
2	(501.6,511.6,521.6)	(0.8,0.85,0.95)	(0.75,0.85,0.95)	(0.75,0.85,0.95)	(0.9,0.9,0.9)
3	(627,637,647)	(0.7,0.8,0.9)	(0.8,0.9,0.95)	(0.65,0.75,0.85)	(0.9,0.9,0.9)
4	(522.5,532.5,542.5)	(0.7,0.8,0.9)	(0.7,0.8,0.9)	(0.7,0.8,0.9)	(0.9,0.9,0.9)
5	(752.4,762.4,772.4)	(0.8,0.85,0.95)	(0.8,0.9,0.95)	(0.6,0.7,0.8)	(0.9,0.9,0.9)

In order to evaluate the efficiency of the candidate suppliers at the first stage, the five most important qualitative and quantitative criteria, i.e., purchasing price, quality, response time, reliability in delivery, and expiration date, were first identified through interviews with the procurement department's experts. Then, the performances of the suppliers for these criteria were acquired using both available historical data and subjective judgements of the firm's experts and managers. However, as there was uncertainty in the performance of a number of criteria, these values were determined in the form of triangular fuzzy numbers (see Table 22.1).

To compare the overall performance of the suppliers, a constant return to scale DEA model (Charnes et al. 1978) is applied. However, in the presence of uncertainty in a number of parameters (i.e., the inputs and outputs of the DEA model), the proposed DEA model is extended into a fuzzy DEA model.

The notations for the DEA model with crisp data are as follows:

Indices

j Index of DMU, $j \in \{1, 2, \ldots, n\}$
i Index of inputs, $i \in \{1, 2, \ldots, m\}$
r Index of outputs, $r \in \{1, 2, \ldots, s\}$

Parameters

y_{rj} Amount of output r for DMU j
x_{ij} Amount of input i for DMU j

Variables

u_r Weight assigned to output r
v_i Weight assigned to input i

The following model is separately solved for each DMU (i.e., supplier). Any DMU $_j$ under consideration is represented by index o where o ranges over $i = 1, 2, \ldots, n$.

$$\max\ e_o = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \tag{22.20}$$

s.t.

$$\frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \leq 1 \forall j \tag{22.21}$$

$$u_r, v_i \geq 0 \forall i, r \tag{22.22}$$

Objective function (22.20) is trying to optimize the efficiency of the DMU under consideration. Also, Constraint (22.21) states that the efficiency of all DMUs (i.e. Suppliers) should be less than one. The above non-linear model can easily be converted to a linear model, using fractional programming techniques. For this purpose, by defining two variables λ_j and θ_o, the dual linearized model (called envelopment model) is expressed as follows:

$$\min \theta_o \tag{22.23}$$

s.t.

$$\theta_o x_{io} - \sum_{j=1}^{n} \lambda_j x_{ij} \geq 0 \forall i \tag{22.24}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro} \forall r \tag{22.25}$$

$$\lambda_j \geq 0 \forall j, \theta_o \text{free} \tag{22.26}$$

Now, by defining y_{rj} and x_{ij} as triangular fuzzy parameter as $\tilde{y}_{rj} = \left(y_{rj}^1, y_{rj}^2, y_{rj}^3\right)$ and $\tilde{x}_{ij} = \left(x_{ij}^1, x_{ij}^2, x_{ij}^3\right)$, the fuzzified DEA model can be rewritten as follows:

Table 22.2 Efficiency obtained for each supplier using the DEA model

Supplier	1	2	3	4	5
Efficiency	0.448	1	0.945	1	0.789

$$\min \theta_o \tag{22.27}$$

s.t.

$$\theta_o \tilde{x}_{io} - \sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} \geq 0 \forall i \tag{22.28}$$

$$\sum_{j=1}^{n} \lambda_j \tilde{y}_{rj} \geq \tilde{y}_{ro} \forall r \tag{22.29}$$

$$\lambda_j \geq 0 \forall j, \tilde{\theta}_o \text{free} \tag{22.30}$$

In order to be able to deploy the above developed DEA model, the first criterion (i.e., purchasing price) is considered as input and the remainder criteria as outputs. The resulted efficiencies for different suppliers are reported in Table 22.2.

As can be seen from Table 22.2, suppliers #2 and #4 have taken the highest efficiency rates and supplier #1 has the worst efficiency among the evaluated suppliers. Hence, at the current situation, it can be recommended to managers to focus more on establishing relationships with suppliers #2 and #4. However, it does not mean that the other suppliers would not be used in the master plan. The scores obtained in Table 22.2 are employed as input of the second stage of the framework where the supplier selection and order allocation problem is solved.

22.5 Conclusions

This chapter investigated the procurement management in healthcare systems from the perspective of operations management. In this regard, the growing importance of healthcare services in today's world was first presented. Different definitions in the context of procurement management were investigated and the factors affecting them were introduced and discussed. Then, the importance and advantages of procurement management in the healthcare sector, which justify the motivation of preparing the chapter, were presented. The main characteristics of procurement which influence healthcare systems and the related supply chains were also examined.

Investigating and dealing with the procurement management related decision problems in the healthcare systems using the operations research approaches constitutes the main part of the chapter. The relevant problems were examined under the main subjects of planning in GPOs, supplier selection, order splitting,

bundling products, e-procurement and performance measurement. In addition, to provide a perfect understanding about the presented material, some examples and also a practical case study were provided.

As utilizing the operational research approaches (e.g. optimization techniques) in the context of healthcare systems is an emerging research area compared to industrial and some other service sectors, this area is faced with many challenges and great research gaps, which could be approached by academics and practitioners. This issue is more evident in the procurement management area, as we perceived throughout the chapter.

As the final point, it is noteworthy that the procurement management has relatively a rich literature in industrial and manufacturing environments, while in the healthcare setting, great gaps are observed. Nevertheless, although the healthcare sector has the specific features, which differentiate it from other industries, but many ideas and solution methods presented in the commercial procurement management can be applied to healthcare systems.

References

Adobor H, McMullen RS (2014) Strategic purchasing and supplier partnerships—the role of a third party organization. J Purch Supply Manag 20(4):263–272

Aissaoui N, Haouari M, Hassini E (2007) Supplier selection and order lot sizing modeling: a review. Comput Oper Res 34(12):3516–3540

Ammenwerth E, Gräber S, Herrmann G, Bürkle T, König J (2003) Evaluation of health information systems—problems and challenges. Int J Med Inform 71(2):125–135

Baier C, Hartmann E, Moser R (2008) Strategic alignment and purchasing efficacy: an exploratory analysis of their impact on financial performance. J Supply Chain Manag 44(4):36–52

Bakker E, Walker H, Schotanus F, Harland C (2008) Choosing an organisational form: the case of collaborative procurement initiatives. Int J Procure Manag 1(3):297–317

Baltacioglu T, Ada E, Kaplan MD, Yurt And O, Cem Kaplan Y (2007) A new framework for service supply chains. Serv Ind J 27(2):105–124

Beil DR (2009) Supplier selection. Stephen M. Ross School of Business, Ann Arbor

Benyoucef M, Canbolat M (2007) Fuzzy AHP-based supplier selection in e-procurement. Int J Serv Oper Manag 3(2):172–192

Beşkese A, Evecen C (2012) Supplier selection in healthcare sector. J Trends Dev Mach 16(1): 91–94

Blair RD, Durrance CP (2013) Group purchasing organizations, monopsony, and antitrust policy. Manag Decis Econ 35:433–443

Burns L (2002) The health care value chain: producers, purchasers, and providers. Jossey-Bass, San Francisco

Burns LR (2014) The performance of Group Purchasing Organizations (GPOs) in the health care value chain: a literature review. Department of Health Care Management, The Wharton School, University of Pennsylvania, Philadelphia

Burns LR, Lee JA (2008) Hospital purchasing alliances: utilization, services, and performance. Health Care Manag Rev 33(3):203–215

Carr AS, Smeltzer LR (1997) An empirically based operational definition of strategic purchasing. Eur J Purch Supply Manag 3(4):199–207

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444

Chen IJ, Paulraj A, Lado AA (2004) Strategic purchasing, supply management, and firm performance. J Oper Manag 22(5):505–523
Cunningham W, Wilson J (1992) Stockless inventory systems for the health care provider: three successful applications. J Health Care Mark 12(2):39–45
Dalkilic V, Lodree Jr EJ, Ramesh A, Sarpkaya G (2006) Order splitting for multiple-supplier multiple-item healthcare supply chains. Paper presented at the IIE annual conference, Florida, 20–24 May 2006
De Boer L (1998) Operations research in support of purchasing: design of a toolbox for supplier selection. Twente University Press, Netherlands
De Boer L, Labro E, Morlacchi P (2001) A review of methods supporting supplier selection. Eur J Purch Supply Manag 7(2):75–89
DeJohn P (1998) Columbia's ABC program: will it die a natural death? Hosp Mater [Dollar Sign] Manag 23(5):1–9
Duclos LK (1993) Hospital inventory management for emergency demand. Int J Purch Mater Manag 29(3):29–38
Essig M, Arnold U (2001) Electronic procurement in supply chain management: an information economics-based analysis of electronic markets. J Supply Chain Manag 37(3):43–49
Fadlalla A, Wickramasinghe N (2004) An integrative framework for HIPAA-compliant I* IQ healthcare information systems. Int J Health Care Qual Assur 17(2):65–74
Fearon HE, Dobler DW, Killen KH (1992) The purchasing handbook. McGraw-Hill Companies, New York City
Figueras J, Robinson R, Jakubowski E (2005) Purchasing to improve health systems performance. McGraw-Hill Education, UK
Ford EW, Hughes JA (2007) A collaborative product commerce approach to value-based health plan purchasing. Supply Chain Manag Int J 12(1):32–41
Ford EW, Scanlon DP (2007) Promise and problems with supply chain management approaches to health care purchasing. Health Care Manag Rev 32(3):192–202
Gary Jarrett P (1998) Logistics in the health care industry. Int J Phys Distrib Logist Manag 28(9/10):741–772
Government US (1996) National communications system technology and standards division. General Services Administration Information Technology Service, Standard Glossary of Telecommunications Terms, USG, Washington, DC
Gurgur C (2013) Healthcare product procurement in dual supplied systems. Paper presented at the 7th IFAC conference on manufacturing modelling, management, and control, Saint Petersburg, 19–21 June 2013
Ho W, Xu X, Dey PK (2010) Multi-criteria decision making approaches for supplier evaluation and selection: a literature review. Eur J Oper Res 202(1):16–24
HSCA (2011a) A primer on group purchasing organizations: questions and answers. Healthcare Supply Chain Association (HSCA), Washington, DC
HSCA (2011b) Understanding how group purchasing organizations work. Healthcare Supply Chain Association, Washington, DC
Hu Q, Schwarz LB, Uhan NA (2012) The impact of group purchasing organizations on healthcare-product supply chains. Manuf Serv Oper Manag 14(1):7–23
Hu QJ, Schwarz LB (2011) Controversial role of GPOs in healthcare-product supply chains. Prod Oper Manag 20(1):1–15
Issel LM (2005) Straddling attention to a healthy bottom line and health effects. Health Care Manag Rev 30(4):281
Kalantari M (2015) A robust master planning approach to drug supply chain under disruption: a case study on ATRA Pharmaceutical Company (in Persian). MSc thesis, Iran University of Science and Technology, Tehran
Kameshwaran S, Narahari Y (2003) e-Procurement using goal programming. E-commerce and web technologies. Springer, Prague, pp 6–15
Ketikidis PH, Kontogeorgis A, Stalidis G, Kaggelides K (2010) Applying e-procurement system in the healthcare: the EPOS paradigm. Int J Syst Sci 41(3):281–299

Ketikidis PH, Lenny Koh S, Gunasekaran A, Danas K, Roudsari A, Ketikidis PH (2006) The applicability of a multi-attribute classification framework in the healthcare industry. J Manuf Technol Manag 17(6):772–785
Kirsch G (2002) The business of e-Health. Int J Med Mark 2:106–111
Klein R (2012) Assimilation of internet-based purchasing applications within medical practices. Inf Manag 49(3):135–141
Knudsen D (1999) Procurement performance measurement system. Dissertation. Lund University, Lund
Kumar A, Ozdamar L, Peng Ng C (2005) Procurement performance measurement system in the health care industry. Int J Health Care Qual Assur 18(2):152–166
Lambert DM, Adams RJ, Emmelhainz MA (1997) Supplier selection criteria in the healthcare industry: a comparison of importance and performance. Int J Purch Mater Manag 33(4):16–22
Leenders MR, Fearon HE, Flynn A, Johnson PF (2001) Purchasing and supply management. McGraw-Hill College, New York City
Lehmann D, O'Shaughnessy J (1974) Difference in attribute Importance for different industrial products. J Mark 38(2):36–42
Littlejohns P, Wyatt JC, Garvican L (2003) Evaluating computerised health information systems: hard lessons still to be learnt. BMJ 326(7394):860–863
Lysons K, Farrington B (2006) Purchasing and supply chain management. Pearson Education, New York City
Marvel HP, Yang H (2008) Group purchasing, nonlinear tariffs, and oligopoly. Int J Ind Organ 26(5):1090–1105
McCormick RE, Shughart WF, Tollison RD (2006) A theory of commodity bundling in final product markets: Professor Hirshleifer meets Professor Becker. Int Rev Law Econ 26(2):162–179
Minner S (2003) Multiple-supplier inventory models in supply chain management: a review. Int J Prod Econ 81:265–279
Monczka R, Handfield R, Giunipero L, Patterson J (2008) Purchasing and supply chain management. Cengage Learning, Boston
Mousazadeh M, Torabi S, Zahiri B (2015) A robust possibilistic programming approach for pharmaceutical supply chain network design. Comput Chem Eng 82:115–128
Murthy NN, Soni S, Ghosh S (2004) A framework for facilitating sourcing and allocation decisions for make-to-order items. Decis Sci 35(4):609–637
Nagarajan M, Sošic G, Zhang H (2009) Stable group purchasing organizations. Working paper. University of British Columbia, Vancouver
Nollet J, Beaulieu M (2003) The development of group purchasing: an empirical study in the healthcare sector. J Purch Supply Manag 9(1):3–10
Okoroh M, Gombera P, John E, Wagstaff M (2001) Adding value to the healthcare sector-a facilities management partnering arrangement case study. Facilities 19(3/4):157–164
Papageorgiou LG (2009) Supply chain optimisation for the process industries: advances and opportunities. Comput Chem Eng 33(12):1931–1938
Pohl M, Förstl K (2011) Achieving purchasing competence through purchasing performance measurement system design—a multiple-case study analysis. J Purch Supply Manag 17(4):231–245
Pooler VH, Pooler DJ (1997) Measurements of purchasing performance. In: Purchasing and supply management. Springer US, Boston, MA, pp 335–358
Presutti WD (2003) Supply management and e-procurement: creating value added in the supply chain. Ind Mark Manag 32(3):219–226
Qi X (2007) Order splitting with multiple capacitated suppliers. Eur J Oper Res 178(2):421–432
Rao R, Peter M, David S (2006) Procurement strategies for health commodities: an examination of options and mechanisms within the commodity security context. USAID, Arlington
Rego N, Claro J, de Sousa JP (2014) A hybrid approach for integrated healthcare cooperative purchasing and supply chain configuration. Health Care Manag Sci 17(4):303–320
Rosenthal EC, Zydiak JL, Chaudhry SS (1995) Vendor selection with bundling. Decis Sci 26(1):35–48

Ross AD, Jayaraman V (2009) Strategic purchases of bundled products in a health care supply chain environment. Decis Sci 40(2):269–293
Saha RL, Seidmann A, Tilson V (2010) A research agenda for emerging roles of healthcare GPOs and their evolution from group purchasing to information sharing to strategic consulting. Paper presented at the 43rd Hawaii International Conference on System Sciences (HICSS), Honolulu, 5–8 Jan 2010
Saha RL, Seidmann A, Tilson V (2011) Unexpected motivations behind Joining Group Purchasing Organization (GPO). Paper presented at the 44th Hawaii International Conference on System Sciences (HICSS), Kauai, 4–7 Jan 2011
Saltman RB, Figueras J, Sakellarides C (1998) Critical challenges for health care reform in Europe. McGraw-Hill Education, New York City
Schneller ES (2009) The value of group purchasing-2009: meeting the needs for strategic savings. Health care sector advances. Health Care Sector Advances, Arizona
Schneller ES, Smeltzer LR (2006) Strategic management of the health care supply chain. Jossey-Bass, San Francisco
Schotanus F, Telgen J, de Boer L (2010) Critical success factors for managing purchasing groups. J Purch Supply Manag 16(1):51–60
Serb C (2010) Does your GPO deliver the goods? Hosp Health Netw/AHA 84(7):38–40. 42
Silva CA, Sousa J, Runkler TA, JSD C (2006) Distributed optimisation of a logistic system and its suppliers using ant colonies. Int J Syst Sci 37(8):503–512
Simonsen L, Kane A, Lloyd J, Zaffran M, Kane M (1999) Unsafe injections in the developing world and transmission of bloodborne pathogens: a review. Bull World Health Organ 77(10):789–800
Smith AD, Correa J (2005) Value-added benefits of technology: e-procurement and e-commerce related to the health care industry. Int J Health Care Qual Assur 18(6):458–473
Songhori MJ, Tavana M, Azadeh A, Khakbaz MH (2011) A supplier selection and order allocation model with multiple transportation alternatives. Int J Adv Manuf Technol 52(1–4):365–376
Stentoft Arlbjørn J, Pazirandeh A (2011) Sourcing in global health supply chains for developing countries: literature review and a decision making framework. Int J Phys Distrib Logist Manag 41(4):364–384
Swift CO (1995) Preferences for single sourcing and supplier selection criteria. J Bus Res 32(2):105–111
Talluri S, Chung W, Narasimhan R (2006) An optimization model for phased supplier integration into e-procurement systems. IIE Trans 38(5):389–399
Torabi S, Baghersad M, Mansouri S (2015) Resilient supplier selection and order allocation under operational and disruption risks. Transp Res Part E: Logist Transp Rev 79:22–48
Upadhaya B, Munir R, Blount Y (2014) Association between performance measurement systems and organisational effectiveness. Int J Oper Prod Manag 34(7):853–875
Van Buer MG, Venta ER, Zydiak JL (1997) Heuristic approaches to purchasing bundles of multiple products from multiple vendors. J Bus Logist 18(1):125
Van Weele AJ (2009) Purchasing and supply chain management: analysis, strategy, planning and practice. Cengage Learning EMEA, Andover
WHO (2000) World health report. World Health Organiztion (WHO), Geneva
Wikipedia (2015) Performance measurement. Wikipedia, the free encyclopedia

Index

C. Kahraman, Y.I. Topcu (eds.), *Operations Research Applications in Health Care Management*, International Series in Operations Research & Management Science 262, https://doi.org/10.1007/978-3-319-65455-3