Flex Serv Manuf J (2016) 28:526–549
DOI 10.1007/s10696-015-9231-5

Daily nurse requirements planning based on simulation of patient flows

Debora Sarno[1] · **Maria Elena Nenni**[2]

Published online: 23 October 2015

Abstract Nurses account for approximately 50 % of total hospital budgets and their allocation to medical units and shifts can significantly affect the quality of care provided to patients. The adoption of flexible shift schedules and the assessment of actual nursing time can enable sensible resource planning, balancing the quality of care with efficiency in resource use. Starting from the concept that nurse requirements are triggered by patient needs, which are stochastic in nature both for clinical activities and their duration, this paper proposes an innovative Nurse Requirement Planning model grounded on the concept of the clinical pathway (the "standard" sequence of diagnostic, therapeutic and care activities a patient with certain pathology should undertake over time) with its inner routing probability and patient dependence on nurses, which can be correlated to the time needed to perform nursing tasks. In merging and modelling these two aspects, the method summarizes the best features of acuity-quality and timed-task/activity techniques, well known although not usually applied for reasons of demands on clinicians' time. Instead, in this paper, for each shift of the day, hospital management is enabled to choose the optimal number of nurses to meet actual requirements according to a desired service level and personnel saturation by means of a tool that simulates the patient flow in a medical unit based on automatic data retrieval from hospital databases. The validation and verification of the proposal were undertaken in a stroke unit.

Keywords Nurse requirement planning · Clinical pathway · Patient dependency · Healthcare operations management · Patient flow simulation · Stroke unit

✉ Debora Sarno
desarno@unisa.it

1 Deparment of Industrial Engineering, University of Salerno, Fisciano, Italy

2 Deparment of Industrial Engineering, University of Naples Federico II, Naples, Italy

1 Introduction

Management decisions concerning healthcare operations are influenced by conflicting objectives, such as quality of care improvements, cost minimization, resource usage maximization, management of an increasing number of patients within a limited time span and budgetary and multi-facility management in a single location (Bhattacharjee and Ray 2014). The managerial challenge is then to find a trade-off between service offered to patients and efficiency for providers (Brailsford and Vissers 2011), organizing the available resources in a cost-effective way (Grieve et al. 2001).

Personnel are the largest component of hospital costs and a significant share is due to nurses (Massachusetts Hospital Association 2010), whose allocation to medical units significantly (around 50 %) affects budgets (Naidu et al. 2000). On the other hand, unbalanced or understaffed nursing teams can influence the quality of care (Hurst 2003), even determining burnouts and job dissatisfaction for nurses and a higher risk of mortality for patients (Aiken et al. 2002). Predicting workload over time and allocating nurses accordingly are essential for guaranteeing quality of care in a cost-effective manner (Kortbeek et al. 2015; Molema et al. 2007). This can be realized by means of a flexible workforce, which is considered a powerful tool to respond to variability in patient demand by many authors (Kortbeek et al. 2015). It has been adopted all over the world in various modes, including part-time employees, overtime, temporary agency employees and float nurses (Gnanlet and Gilland 2009), providing pools of cross-trained nurses assigned to specific care units at the start of each shift, in an attempt to balance nurse requirement fluctuations among wards while reducing the total employee buffer capacity.

Nurse requirements are triggered by patient needs over their stay in the hospital (patient flow), which are stochastic in nature in terms of clinical activities and their duration, due to pathology, status and system uncertainty (Harper 2002; Bhattacharjee and Ray 2014). Moreover, the patient flow depends on the clinical pathway (Bhattacharjee and Ray 2014), the "standard" sequence of diagnostic, therapeutic and care activities a patient with certain pathology and according to medical guidelines and resource availability should undertake in a hospital. The concept of the clinical pathway was introduced in the UK in 1985 to preserve cost effectiveness and quality of care (Zander 2002) and it is a method for providing care to a defined group of patients over a defined period of time based on guidelines, best practice and evidence-based medicine (Cardoen and Demeulemeester 2008). By this means, it is possible to ease communication and coordination among the different actors involved (physicians, nurses, patients and their relatives, etc.), sequencing the tasks to be carried out and allowing the planning of resources and outcomes (Cardoen and Demeulemeester 2008).

Nowadays, there is an emergent trend in the literature for estimating the routing of patient flows (Adeyemi et al. 2010) through the use of statistical or empirical modelling methods (also financed by national governments and international institutions) based on the analysis of Electronic Medical Records (EMRs), hospital information databases and—more generally—big data. Moreover, many studies have employed clinical data to provide better management of hospitals' resources and materials [for example in the field of drug management; see Iannone et al. (2011, 2014), Guida et al. (2012), Iannone et al. (2013, 2015)].

This paper presents a new model for selecting the appropriate number of nurses per shift based on the probabilistic service level of care. The model is based on clinical flows, which are stochastic both in terms of tasks to be performed and the duration of single tasks, dependent on the acuity of patients [expressible, in the case of nurses' activities, by the dependence of a patient on nurses (Kaliszer 1976)].

The paper is structured as follows: Sect. 2 presents the four-step nurse requirement planning policy, a review of the literature on Nurse Requirement Planning (NRP) and the innovative contribution of the proposed method; Sect. 3 addresses the modelling of the proposed method; Sect. 4 describes the case study chosen to assess the usability of the proposal, that is a stroke unit, setting out the data collection and analysis, simulation, validation and verification. The conclusions follow in Sect. 5.

2 Nurse requirement planning

2.1 Four-stage requirement planning

The issue of Nurse Requirements Planning (NRP) can be contextualized in the wider framework of "nurse capacity planning and scheduling" (Punnakitikashem et al. 2008; Siferd and Benton 1994) as described below (Fig. 1).

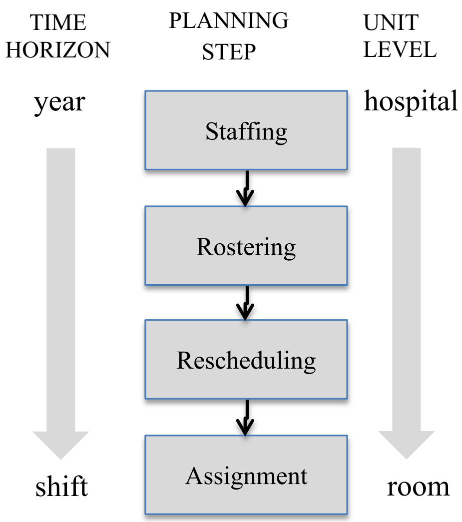

Fig. 1 Framework for planning and scheduling nurse requirements

2.1.1 Nurse staffing

Over the long-term planning horizon (1 or more years), on the basis of the forecasting of patient access and nurse workload, decisions are made about the number and the mix of nurses needed to meet patient demand. This can be made at the departmental level (Jelinek and Kavois 1992), defining the medical unit budget, or at the hospital level (Siferd and Benton 1994), postponing the attribution to a specific unit.

2.1.2 Nurse scheduling (or rostering)

Over the medium-term planning horizon (several weeks), the allocation of nurses to shifts is made taking into account the following constraints (Naidu et al. 2000; Ernst et al. 2004; Clark and Walker 2011):

- Equal distribution of night and weekend shifts among nurses;
- Consideration of nurse preferences;
- Balancing of skilled and unskilled nurses for each shift;
- Clinical requirements based on patient census and acuity;
- Ergonomic evaluations (effect of shifts on people's well-being in terms of the resetting of the biological clock and exposure to variable work schedules).

2.1.3 Nurse rescheduling

Some time (hours) before each shift (for example, in Siferd and Benton (1994), at least one shift before, during a conference call or a daily meeting among nurse supervisors), the number and the attribution of individual nurses to a shift can be reviewed. This can be made on the basis of expected patients' number and characteristics by means of patient census, information concerning Admission, Discharge and Transfer (ADT) (Bard and Purnomo 2004) and actual patient needs (Siferd and Benton 1994), but also nurse absenteeism.

In the case of undersizing, part-time nurses, off-duty nurses (regular staff) or outside nurses (Bard and Purnomo 2004), that is agency nurses, casuals or floaters (nurses employed by the hospital but not assigned to a specific unit), can be called on to meet the expected patient needs (Punnakitikashem et al. 2008; Clark and Walker 2011). In contrast, in the case of oversizing, nurses can be floated to another unit, reassigned to the following day, or put on an on-call list (Bard and Purnomo 2004). There is a substantial body of literature on the issue of rescheduling while keeping the disruption to staff to the minimum (Bard and Purnomo 2004) because meeting nurse preferences has a positive effect on nurses' behaviour (Clark and Walker 2011). Several negotiation algorithms for distributed nurse rostering and rescheduling have been developed (Burke and Curtois 2014).

2.1.4 Nurse assignment

At the beginning of a shift, the patients are assigned to specific nurses, bearing in mind the uncertainty of patient care, nurses' skills and load balancing (Punnakitikashem et al. 2008) and patient location (for example, rooms).

2.2 Methods for estimating nurse requirements

Methods to determine nurse requirements (from staffing to rescheduling issues) have been developed since the 1980s and range from consensus approaches to top-down management approaches (Naidu et al. 2000). They can be grouped into five categories ordered from the simplest to the most complex as described below (Hurst 2003).

2.2.1 Expert judgment

Using expert judgment, the number of nurses per shift is set by managers. To size the staff over 1 year, the duration of each shift, the overlap between shifts, the percentage of personnel time-out and the number of weekly shifts per nurse have to be defined. This approach is evidently subjective and it is difficult to assess the nursing quality or the load variation depending on the patient case mix, but it can easily be applied and allows triangulation with the results of other methods.

2.2.2 Number of nurses per occupied bed

The model considers statistics regarding the number of nurses attributed to the beds of a certain ward, characterized by the primary treatment specialty (elderly, surgical, ophthalmic, etc.) and having passed a quality test. Moreover, nurses are classified by grade, which enables grade mix assessments. The method can be affected by the ward sample and no attention is paid to the nurses' tasks on the ward (surgical support in operating rooms, etc.) or patient dependence on nurses.

For example, Ghosh and Cruz (2005) presented a simple mathematical model to estimate nurse requirements in a hospital, distinguishing among non-critical inpatients, critical inpatients, outpatients and other patients (e.g. surgical). Nurses were assigned to four different care units characterized by distinct load parameters, such as the number of beds available, as well as statistics related to occupation or the available number of working days, the duration of shifts, etc. Another recent contribution is that of Kortbeek et al. (2015), addressing a stochastic method for allocating nurse staffing based on hourly bed census prediction, but not considering the specific health condition (acuity) and needs of patients occupying beds.

2.2.3 Acuity-quality methods

The main hypothesis of acuity-quality models is that the time needed to give assistance to patients is proportional to their dependence on nurses (expressed by a patient dependency category) or patient acuity (Ernst et al. 2004) and changes from ward to ward. Indirect care is considered proportional to direct care and meal breaks and time-out are added to the total direct care for the patient mix to establish the total nursing workload for a period of time. Again, the nurse grade involvement can be taken into account. This kind of model is affected by workload estimation for each patient category, but does not consider individual patient needs.

The classification of the patients is clearly one of the main issues in these methods (Hurst 2005) and many clustering techniques have been proposed. For example, some authors (Needleman et al. 2002) have proposed the use Diagnosis Related Groups (DRGs), widely employed for service reimbursement reasons (Mathauer and Wittenbecher 2013) and based on the concept that some pathologies can require the same amount of resources for nursing staff purposes. Other studies have reported on the use of patient points systems and patient acuity judgments (Brander and Norton 1991).

An extensive survey of English stroke units (Rudd et al. 2009) was carried out in 2006 with the aim of identifying the most relevant patient factor influencing nurse staffing levels. It was found that patient dependency, expressed by the Barthel scale value, has a correlation with nurses' direct workload (Spearman's correlation coefficient of -0.5).

An interesting example of day-to-day nurse scheduling based on patient acuity was reported by Siferd and Benton (1994), who, given a mean patient acuity for all patients expressed in terms of the number of nurses required, considered a mean rate of change in patient condition over the shift, assuming that new patients have higher acuity than those who are hospitalized. Then, they carried out a number of simulations to assess the impact of size of unit, acuity and number of admissions and discharges on the number of nurses needed.

Another case comes from Liang and Turkcan (2015), who presented a multi-objective optimization model for determining the number of nurses and the scheduling of patients (minimizing waiting time) in an outpatient oncology clinic. The sum of patient acuities was matched with the maximum acuity manageable by a nurse to make assignments. A nurse's direct workload was concentrated in the first 30 min after each patient treatment start time (later, the nurse had to monitor contemporaneous patient infusions). Both acuities and nurse workload for the treatment (excluding the other activities of the patient pathway) were assumed to be deterministic.

2.2.4 Timed-task/activity methods

The timed-task/activity approach is based on the concept that the type and frequency of patient interventions are good predictors of nursing time. To apply the methods it is necessary to define the duration for each possible activity and then construct the care plan for each patient. The ward overhead time has to be added to the total direct care time. The main obstacle to the spreading of such an approach is the effort needed to keep detailed care plans for each patient.

Contributing to the debate concerning the use of patient acuity and task-oriented methods to evaluate nurse requirements, a comparative case study was carried out in three medical units examining the CASH method (developed by the Commission for

Administration Services in Hospital) and the GRASP method (designed by the Grace Hospital in Morgantown), which showed that load forecasts can range from 5 to 20 % (Schroeder et al. 1984). The authors concluded that the category system is preferable to the task-based method, which requires much more detailed information concerning patient activities.

2.2.5 Statistics-based methods

Finally, some authors have proposed regression methods, which do not rely on demand-side nurse staff planning. The models are highly dependent on the context in which they are applied, involve a number of independent variables that are sometimes difficult to express quantitatively and cannot be extended beyond the observed range of the variables' value, even if the aggregated data used as inputs for the calculation are usually easy to collect (Hurst 2003).

2.3 Proposed method and innovative contribution

This paper proposes a day-to-day NRP method based on the tasks involved in the patient clinical pathway and patient dependence on nurses in the activities of daily living, correlated to the time needed to perform each nursing task. With these premises, the method overcomes: (i) the limits of timed-task/activity approaches, because it does not require any activity checklist completion by nurses (the information is provided by the clinical pathway) and takes into account the specific conditions of patients (thanks to the patient dependence quantification); (ii) the inaccuracies of acuity quality methods, avoiding the necessity of considering the average value of nursing time required for each patient but rather focusing on single clinical needs.

Moreover, because the aforementioned concepts involve probabilistic evaluations related to task occurrence and duration for each patient, taking into account that patients' number and characteristics also change over a day due to new admissions, discharges and variation in status, the nursing time and consequently the number of nurses required to meet patient demand demonstrate stochastic behaviour. Then, the service level offered to patients (that is the probability of meeting total patient demand) is stochastically dependent on the number of nurses assigned.

Numerical simulation has been used to simulate each patient flow and find the cumulated probability distribution function of the number of nurses required for each shift. This distribution can then be used by management to choose the optimal number of nurses for real patient requirements according to the desired service level and personnel saturation.

The main innovative contributions of this paper in the NRP literature and practice are as follows:

- The proposal of a model which takes into account both the clinical tasks to be performed (because each single task in the clinical pathway and the related probability of occurrence are considered) and the acuity of patients (because patient dependency is considered). In doing so, the model incorporates the main features of timed-task/activity and acuity-quality methods, accurately reproducing the effective nurse requirements based on the clinical tasks to be performed for each patient without neglecting the time needed to carry these out in relation to the acuity of the patient.

- Probabilistic assessment of the service level offered to patients depending on the number of nurses assigned for each shift and related staff saturation.
- Negating the need for time-consuming data entry for the NRP calculation (which often prevents the implementation of other models) through the automatic retrieval of data from EMRs and database of clinical pathways in the hospital's Information Systems.

The model has been validated and verified in a real case study of a stroke unit.

3 Modelling and simulation

In this section, the notation (3.1) and formalization (3.2) are first introduced. Then, the simulation model is presented (3.3).

3.1 Notation

3.1.1 Indices

d: day of the clinical pathway
e: emergency activity of the clinical pathway
o: ordinary activity of the clinical pathway
p: patient
s: nurse shift, with $s = \begin{cases} 1, & morning \\ 2, & afternoon \\ 3, & night \end{cases}$

3.1.2 Parameters

$\dot{b}_p$: Barthel scale value assigned to patient p for his/her next day of the pathway $\dot{b}_p \in [0, 20]$
$\gamma_{s,p}$: percentage of shift time for emergency activities related to patient p
φ_e and $\varphi_{o,d,s}$: probability of attribution of activity e or o to nurses (instead of other resources)
h_p: pathway day of patient p
$NICT_s$: nursing indirect care time, i.e. the nursing load to provide indirect care in the medical unit during s
m_e and $m_{o,d,s}$: minimum duration of activities e and o

$\bar{m}_e$ and $\bar{m}_{o,d,s}$: average duration of activities e and o
M_e and $M_{o,d,s}$: maximum duration of activities e and o
ρ_e and $\rho_{o,d,s}$: probability of occurrence of activities e and o
$\pi_{s,p}$: percentage of duration of emergency activities related to patient p over shift s
t_s: duration of the shift
v: maximum availability of a nurse during a shift (percentage)
$w_{s,p}$: presence of patient p on the ward during shift s (to take into account discharge or transfer) with $w_{s,p} = \begin{cases} 1 & \textit{if patient p is present during s} \\ 0 & \textit{otherwise} \end{cases}$
x: uncertainty in the attribution of the Barthel scale value

3.1.3 Variables

$NDCT_s$: nursing direct care time, i.e. nursing loading to provide direct care to patients in the medical unit during s
n_s: number of nurses required during shift s
$RO_{s,p}$: nursing time requirement for ordinary activities during shift s for patient p
$RE_{s,p}$: nursing time requirement for emergency activities during shift s for patient p
$t_{e,p}$: duration of activity e assigned to patient p
$t_{o,d,s,p}$: duration of activity o, referred to as day d and shift s of the pathway, assigned to patient p
TNT_s: total nursing time in the medical unit during s

3.2 System formalization

The computation of the number of nurses n_s required to answer the total demand for a shift s is gradually described using a top-down approach.

If the maximum availability of nurses (v) is considered to exclude breaks for physiological reasons or inefficiencies in the healthcare organization in terms of communications between actors from working time [the practical capacity, as defined by Kaplan and Anderson (2004)], the number of nurses n_s required to fulfil all the requirements of a shift s is equal to the total nursing time (TNT_s) divided by the shift duration t_s and the maximum availability of nurses v(1):

$$n_s = \left\lceil \frac{TNT_s}{v * t_s} \right\rceil \tag{1}$$

The total nursing time (TNT_s) is due to nursing time for direct care activities $NDCT_s$ (from the planning of nursing activities to meal assistance) and the indirect care activities $NICT_s$ (for example, drug management), independent of the number of patients hospitalized but dependent on the size of the medical unit and its specialties:

$$TNT_s = NDCT_s + NICT_s \tag{2}$$

The activities of the clinical pathway are classified as “emergency”, e, performed to stabilize a patient coming from the emergency department and “ordinary”, o, or ADL.

Then, the nursing time for direct care over each shift s ($NDCT_s$) is expressed as the sum of the nursing time required by all patients p to perform emergency and ordinary activities ($RE_{s,p}$, $RO_{s,p}$) in their pathways:

$$NDCT_s = \sum_{p=1}^{P} (RO_{s,p} + RE_{s,p}) \tag{3}$$

The two terms of (3) are computed by summing up the duration of single patient activities, making the following assumptions. Each activity has a minimum m, average $\bar{m}$ and maximum M time needed for it to be performed. Moreover, each activity has a probability ρ of occurring and a probability φ of being assigned to a nurse rather than to another resource. Finally, activities are sequentially organized over the planning horizon and ordinary activities, which follow emergency activities, are attributed to a single shift s in a hospitalization day d. Then, the duration of each patient’s activity can be found depending on the type (e, o).

3.2.1 Duration of emergency activities

The duration of an emergency activity for a patient p ($t_{e,p}$) is a stochastic variable function of the three duration parameters and the two activity probabilities. Thus:

$$t_{e,p} = f(m_e, \bar{m}_e, M_e, \varphi_e, \rho_e) \tag{4}$$

The emergency phase can last more than one shift. The percentage of the time needed to perform the emergency activities related to patient p over each shift s is expressed by the term $\pi_{s,p}$. An example is shown in Fig. 2.

Then, the total nurse requirement for a patient p ($RE_{s,p}$) can be found:

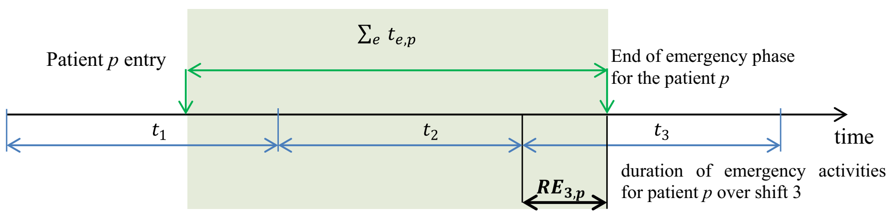

Fig. 2 Example of emergency phase of new patient on the timeline of a day

$$\mathrm{RE_{s,p}} = \pi_{\mathrm{s,p}} * \sum_{\mathrm{e}} \mathrm{t_{e,p}} \tag{5}$$

Finally, the emergency activities related to the patient cover a quoted duration of time in each shift $\gamma_{s,p}$. It is worth reporting the following equivalence:

$$\pi_{s,p} * \sum_{e} t_{e,p} = \gamma_{s,p} * t_s \tag{6}$$

3.2.2 Duration of ordinary activities

The duration of each ordinary activity for a patient p ($t_{o,d,s,p}$) is considered to be a function of the activity duration parameters, the patient dependency and the related uncertainty in the attribution of dependency made by nurses and the two activity probabilities (7).

The dependency value of each patient $\dot{b}_p$ is attributed by nurses using the Barthel scale (Rudd et al. 2009), in which a completely dependent patient has the minimum Barthel value of 0 and an autonomous patient has the maximum Barthel value of 20. Such dependency changes over the days of hospitalization according to the clinical activities in the pathways.

In this paper, according to the correlation found by Rudd et al. (2009), $t_{o,d,s,p}$ is assumed to be a linear function of the Barthel scale (8), with two different slopes of the curve changing in the point of the average values of the variables $\left(\bar{m}_{o,d,s}, 10\right)$ as in Fig. 3.

The attribution of a Barthel value is affected by uncertainty (x) due to the nurses' perception of the patient status, the change in patient status over time and certain inefficiencies in the healthcare system (Bhattacharjee and Ray 2014). Hence:

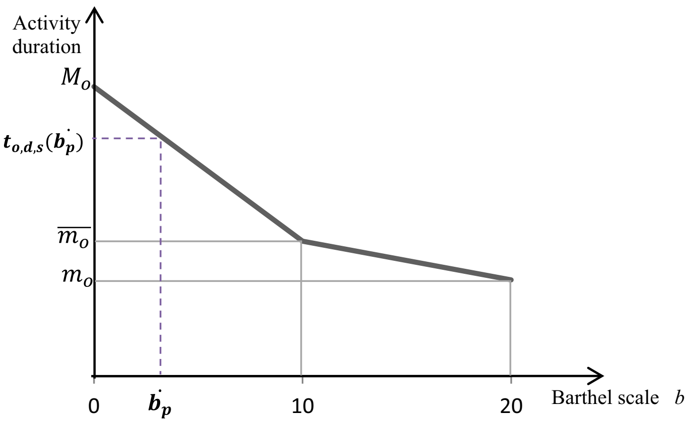

Fig. 3 Activity duration $t_{o,d,s}\left(\dot{b}_p\right)$ in dependence on the Barthel value assigned to a patient

$$t_{o,d,s,p} = f\left(t_{o,d,s}\left(\dot{b}_p\right), x, \varphi_{o,d,s}, \rho_{o,d,s}\right) \quad \text{with} \quad d = h_p \tag{7}$$

$$t_{o,d,s}\left(\dot{b}_p\right) = \bar{m}_{o,d,s} + \frac{\left(\dot{b}_p - 10\right)}{\left(b_{limit} - 10\right)}\left(m_{limit} - \bar{m}_{o,d,s}\right) \tag{8}$$

where: $b_{limit}, m_{limit} = \begin{cases} 0, M_{o,d,s} & 0 \leq \dot{b}_p < 10 \\ 20, m_{o,d,s} & 10 \leq \dot{b}_p \leq 20 \end{cases}$

Finally, the nurse requirement for a patient p ($RO_{s,p}$) is the sum of the durations of the activities planned for the patient ($t_{o,d,s,p}$) according to his/her possible discharge or transfer during the shift ($w_{s,p}$). Moreover, in the case that a patient still has to conclude the emergency phase, the ordinary requirement is a percentage of the total remaining shift time ($1 - \gamma_{s,p}$). Thus:

$$RO_{s,p} = \left(1 - \gamma_{s,p}\right) * \sum_{o,d=h_p} t_{o,d,s,p} * w_{s,p} \tag{9}$$

For the sake of simplicity, the activities analysed are exclusively those carried out by one type of resource (nurse) belonging to one medical unit in which patients follow one clinical pathway. The model can easily be extended to many pathways, many medical units and of course more resources.

3.3 Model development

As recognized by many authors, due to the complexity of the patient pathway, simulation is the best modelling solution to take into account task time, routing probabilities and the integration of facilities (Bhattacharjee and Ray 2014; Cardoen and Demeulemeester 2008). The proposed NRP method is based on a Monte Carlo Simulation (MCS), a very common numeric simulation used to generate randomly a set of events for a stochastic variable according to its Probability Distribution Function (PDF). The choice of a simple MCS rather than a discrete event simulation or other simulation methods is made for two main reasons:

1. Clinical pathways provide the care activities to be performed but no information is given regarding their effective sequence over one shift (for example, patient cleaning may be required before his/her lunch), the execution of more than one contemporaneous activity (for instance, test tube transportation and assistance for lunch may be executed in parallel by two nurses or in sequence by the same one), the exact waiting time for clinical outcomes that can be a trigger for other activities during one shift (for example, only when the blood test results are received, the lunch can take place). Due to this, even if some events happen and are simulated during the shifts, such as arrivals or the occurrence of tasks, to avoid inaccurate evaluations at the hourly level, the durations of tasks are aggregated at the shift level to define the total direct nursing time required in this time bucket.

2. The model has been conceived to be understood, used daily and modified by hospital management. For this reason, it has been developed in a Microsoft Excel Spreadsheet in which MCS can easily be implemented.

By means of MCS, the stochastic behaviour of patient admissions, characteristics, clinical activities and the related duration are reproduced over the shifts to estimate the total nursing time in different scenarios and determine the nursing service level depending on the number of nurses.

3.3.1 Inputs

Some medical unit data are needed:

- Database of pathway activities (e and o), with the duration parameters $m_e, m_{o,d,s}, \bar{m}_{o,d,s}, \bar{m}_e, M_e, M_{o,d,s}$;
- Expected duration of the indirect activities for each shift $NICT_s$;
- Duration t_s and time schedule of shifts;
- Maximum availability of nursing staff v;
- Uncertainty in the attribution of the Barthel scale x;
- PDFs of patient inter-arrival time, duration of emergency activities, Barthel values assigned;
- Probabilities of occurrence ρ and assignment to a nurse φ for each activity.

In daily operations, the following data should be collected for each hospitalized patient p:

- Day of the clinical pathway h_p;
- Possible modification in the pathway (patient flow routing definition);
- Dependence on nurses $\dot{b}_p$;
- Planned discharge or transfer $w_{s,p}$;
- Time of admission to the medical unit.

3.3.2 Simulation of incoming patients

Patients' inter-arrival is simulated based on the related PDF. This time is added to the last admissions to determine the next entries of new patients.

3.3.3 Simulation of patients' requirements

For patients in the emergency phase, the duration of emergency activities e (as described in Sect. 3.2.1) patient dependency $\dot{b}_p$, ρ and φ are simulated. The duration of ordinary activities related to each patient for each shift is simulated (according to the modelling proposed in Sect. 3.2.2), taking into account the Barthel scale value and inner uncertainty, the occurrence of each activity for the patient and the task assignment to nurses.

4 Case study: the stroke unit

To validate and verify the proposed model, it was implemented in a real environment. The case study under analysis is a stroke unit, a particular segment of the neurology medical unit dealing with stroke patients.

Stroke is the second leading cause of disability in Europe and the reason for 10 % of deaths worldwide (Wittenauer and Smit 2013). Its clinical pathway is very expensive for healthcare systems [it is responsible for 7 % of the UK NHS budget (Gillespie et al. 2011)]. It requires the systemic integration of different services, recommended to be organized in networks (Wittenauer and Smit 2013), such as primary care, transfer by ambulance, acute and rehabilitation treatments and care support in the community.

In the hospital environment, due to the high number of stroke cases and the considerable hospital cost, this patient pathway is well known: clinical process flow charts and duration of tasks are standardized and shared among actors (emergency department, medical unit physicians and nurses) to ensure the efficient management of patient care.

As reported by Chemweno et al. (2014), some simulation studies have been developed on stroke patients and stroke units for strategic planning purposes. For instance, Elf et al. (2009) proposed a causal loop diagram of the care process to design a stroke unit; Bayer et al. (2010) presented a Discrete Event Simulation (DES) model of the care pathway from the stroke event to home post-acute care to share the pathway composition among stakeholders and offer management a decision-making tool; the patient flow in an emergency department has been modelled as a decision tree simulated using DES (Chemweno et al. 2014) to propose organizational improvements; the economic convenience of increasing the number of stroke patients receiving the particular and costly treatment of thrombolysis (Gillespie et al. 2011) has been investigated implementing a probabilistic approach to assess the outcomes of such a service and taking into account the drug administration, hospitalization and community care costs.

4.1 Case study data set

The stroke unit under study is within a medium-sized Italian acute university hospital of 1500 beds. It is devoted to the care of ischaemic stroke (classified as DRG 014) and has eight beds for inpatients and one room for incoming cases transferred from the emergency department. The total number of patients treated in the hospital with this DRG in 2014 was 610, with an average length of stay of around 10 days, comprising the first rehabilitation activities of patients. It is clear that the bed capacity of the stroke unit is not sufficient to cater for all stroke patients: 260 patients were admitted to and discharged from the unit in 2014 and many transfers occurred, not only due to lack of available beds but also as a result of clinical complications. This explains the extensive engagement of the ward personnel in the execution of data collection: they daily experience transfers of new patients at any stage of the pathway without having a method to forecast the required nursing time. This research aimed to address such issues and also constituted the starting point for enabling accurate what-if analysis on the possible variations in bed capacity and pathway tasks.

The values of the model inputs presented in Sect. 3.3.1, collected in the stroke unit by means of semi-structured interviews, time measurements and data retrieval from hospital information systems, are listed below, making a distinction between parameters and variability data.

4.1.1 Parameters

- Database of pathway activities (Tables 1 and 2). The definition of the clinical activities over the pathway (5 days) for each shift had been already accomplished by physicians for medical reasons. To share the pathway model with clinicians and enable them to make suggestions to improve its design, a 3D simulation was developed using the software FlexSim Healthcare (Fig. 4).

The three time parameter values of each activity (minimum, maximum and average activity duration) were defined by means of expert judgments. In particular, four persons (physicians and nurses of different grades, depending on the task to be evaluated) were asked to complete a form independently. In the case of divergence of short values (maximum variance-to-mean ratio lower than 10 %), the mean time value of the parameter was taken into account; otherwise, the mean value from 10 time measurements collected during the delivery of care to patients with different acuity was considered. Activity probabilities were collected using the same approach.

- Values of $NICT_s$, t_s, v, x (Table 3). In detail, $NICT_s$ was recorded by nurses for each shift for 10 days. The average value was taken. v was assumed based on the

Table 1 Duration parameters of emergency activities. φ_e and ρ_e are equal to 1

Emergency activities (e)	Duration (min)	
	Min. m_e	Max. M_e
E1	10	30
E2	10	30
E3	10	20
E4	10	20
E5	10	20
E6	10	10
E7	20	50
E8	5	5

Table 2 Attribution of the ordinary activities of pathway to days and shifts, duration parameters and probabilities (in terms of attribution to nurses instead of other personnel and related to the occurrence of the activity for the patient)

Ordinary activities (o)	Day of the pathway d					Shift s			Duration (min)			Probability	
	1	2	3	4	5	1	2	3	Min. $m_{o,d,s}$	Avg. $\bar{m}_{o,d,s}$	Max. $M_{o,d,s}$	Nurse attribution $\varphi_{o,d,s}$ (%)	Activity Occurrence $\rho_{o,d,s}$ (%)
F1	1	1	1	1	1	1			2	5	7	100	100
F2	1	1	1	1	1	1			5	5	5	100	100
F3	1	1	1	1	1	1			2.5	3.5	7.5	100	100
F5	1	1	1	1	1	1			5	10	15	100	100
F6	1	1	1	1	1	1			2.5	10	20	100	100
F7	1					1			50	60	75	100	100
F7 bis		1	1	1	1	1			10	15	20	100	100
F9	1	1	1	1	1	1			2.5	5	7.5	100	20
F11	1	1	1	1	1	1			5	11	15	100	100
F12	1	1	1	1	1	1			8	10	20	100	100
F13	1	1	1	1	1		1		5	5	5	100	100
F14	1	1	1	1	1		1		2.5	5	7.5	100	20
F15	1	1	1	1	1		1		8	10	20	100	100
F16	1	1	1	1	1		1		5	11	15	100	100
F17	1	1	1	1	1		1		5	10	30	100	100
F18	1	1	1	1	1		1		5	5	5	100	100
C5		1				1			50	65	90	35	100
C6			1			1			50	65	90	35	100
C7				1		1			50	65	90	35	35
F20	1					1			5	7	10	100	100
F21		1				1			5	10	20	100	100
A1	1	1	1	1	1			1	10	15	25	100	100
F22		1				1			50	60	80	100	100

reflections of Kaplan and Anderson on the practical capacity of employees (2004). x was obtained after three meetings with clinicians: the first to present the problem and the proposed approach, the second to collect the initial opinions of each individual on its importance and value based on their experience (the Barthel scale was used at the time for each patient for medical purposes) and the third to bring the experts to opinion convergence.

- Daily patient data (Table 4), used for the simulations below, deduced from EMR.
- Data records, such as those in Table 4, and nursing direct care time in each shift ($\mathrm{NDCT_s}$) over 1 month, used to validate the model.

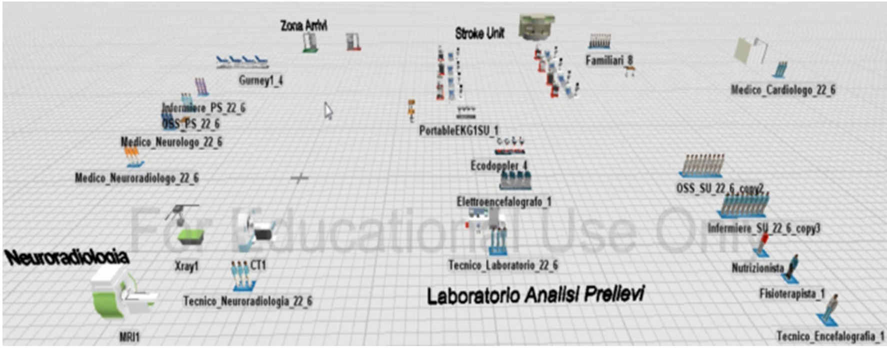

Fig. 4 Stroke patient pathway simulation with FlexSim Healthcare software

Table 3 Other input data

Shift s	$NICT_s$ (min)	t_s (min)
1	60	420
2	120	420
3	30	600
Other input		Value
v		80 %
x		4

Table 4 Patient data for 1 day of the case study—"loaded unit"

Patient p	Barthel scale b_p	Day of the pathway h_p	Time of patient entry	Presence of patient $w_{s,p}$		
				1	2	3
1	0	1	01/01/2015 16:00	1	1	1
2	2	2		1	1	1
3	5	2		1	1	1
4	7	3		1	1	1
5	8	3		1	1	1
6	13	5		1	1	1
7	10	8		1	1	1
8	20	9		1	1	1

4.1.2 Variability data

- PDF of patients' inter-arrival time in the stroke unit, calculated from the EMRs of the patients hospitalized on the ward in 2014, compulsorily recorded for hospital reimbursement purposes in many countries all over the world. SQL queries were used to extract arrival data from hospital databases (both from emergency departments and medical units). Then, the records were cleaned to remove visibly incorrect data, such as entries in which the arrival at the stroke unit happened before arrival at the emergency department. Finally, the data were statistically analysed, showing that the inter-arrival time fitted a negative exponential distribution with a rate parameter of 1.47.

- Duration of emergency activities, assumed to be uniformly distributed between the minimum and the maximum duration $t_e = UNIF(m_e, M_e)$ as nurses only support other clinicians during these tasks and no information on next patients is available.
- Duration of ordinary activities, assumed to be uniformly distributed as $t_{o,d,s,p} = UNIF\left(t_{o,d,s}\left(\dot{b}_p \mp x\right)\right)$. In this way the duration has been proportionally correlated to the Barthel scale but it has been assumed to be uniformly distributed within the variability range determined by the uncertainty x.
- Time series of the Barthel values of the medical unit were not available in the hospital information system because they were recorded daily on paper by physicians (the information was not part of the EMR). To avoid transcription errors on an incomplete sample (many papers relating to the patients' first hospitalization day were missing), the Barthel PDF used to attribute a value to the incoming patient was extrapolated from a previous study (Rudd et al. 2009). Following the study, the information system was updated to collect this information in a structured way.

4.2 Verification, validation and simulation

4.2.1 Verification

The verification phase, that is the process of ensuring that the simulation model would operate as intended, was carried out by means of interviews with clinicians. They checked the model with the authors to ensure that all the important components were included in the model and in the right sequence.

4.2.2 Validation

The validation phase was undertaken to ensure that the simulation model represented the real situation at a given confidence interval. First of all, a table of results (in terms of total nursing time for each shift, duration of emergency activities over the shifts, etc.) was discussed with the same panel of experts to assess the order of magnitude of variables obtained based on their experience (face validity). Then, a statistical validity test was performed to provide a quantitative comparison between the actual system and the simulation model. Thus, the model was validated against real data for the stroke unit. The nursing direct care time for each shift ($NDCT_s$) was collected over a period of 1 month and compared to the simulated data in the same period for the same patients with 1000 replications. Applying in cascade the Chi square procedure for testing normality and the F-test to compare variances, the Smith–Satterthwaite test (which is used when variances are dissimilar) for each descriptive variable reported that no statically significant difference between the real system and the simulated model at the 0.05 level of significance and thus the model is valid.

Finally, a replication analysis was performed to determine the number of replications required to analyse the difference between the alternatives statistically. Specifically, using the set of the number of nurses required (n_s), an absolute error of 10.0 % (that is the percentage amount of dispersion exhibited by the total number of nurses required around its mean divided by the sample mean at a certain confidence level) with a 95 % confidence level, it was found that 1000 replications of each experiment were needed.

4.2.3 Case study simulation

Using the daily patient characteristics from Table 4 and implementing the model in a simple Microsoft Excel Spreadsheet, the cumulated distribution function of the number of nurses required for a shift over 1000 simulation replications is shown in Table 5 (the results for shift 1 are provided in graphical form in Fig. 5a). By this means, the nurse manager is able to make decisions about the number of nurses who might be re-scheduled for the day. For example, for the first shift, four nurses are required to assure a 99.2 % service level to patients, while with three nurses or fewer, the probability of satisfying the patient requirements is equal to 75.1 %. In the second shift, however, the assignment of three nurses or more would be a waste.

Moreover, for each number of nurses per shift, the probability of nurse saturation is provided. As expected, the greater the number of nurses per shift, the lower the average saturation. In the example given in Fig. 5b (shift 1), assigning only three nurses would mean they would be fully loaded (between 70 and 80 % saturation) in 43 % of cases, with a possible negative effect on stress. However, with four nurses, they would be less loaded in the same saturation range (18 %). In all cases, they would be loaded at least at 50 %. In the rescheduling logic, this information should be extremely interesting for the nurse manager in charge of making decisions about nurse recall on duty, overtime or floating.

Table 5 Patient service level depending on the number of nurses required n_s per shift s—"loaded unit"

Number of nurses n_s	Service level achievable with n_s for each s		
	1 (%)	2 (%)	3 (%)
1	0.0	0.0	99.7
2	0.0	98.4	100.0
3	75.1	100.0	100.0
4	99.2	100.0	100.0
5	99.9	100.0	100.0
6	100.0	100.0	100.0

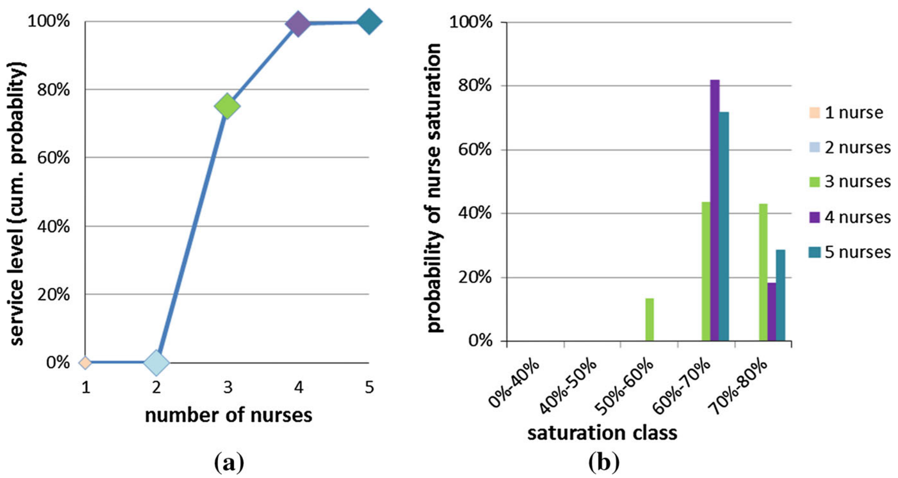

Fig. 5 "Loaded unit" scenario results: **a** patient service level depending on the number of nurses required n_s per shift 1; **b** nurse saturation probability in shift 1

Looking at Table 4, the results provided above were obtained for a fully loaded medical unit (the "loaded unit" scenario) in which: (i) all beds are occupied, (ii) patients are acute (from Barthel scale values), requiring for each activity more than mean nursing time and (iii) patients are in the first days of the pathway, requiring more mean nurse time than in the following days. To show the effectiveness and the usefulness of the approach, another medical unit "load" scenario was tested, in which patients and requirements were smaller ("unloaded unit" scenario), the data for which are shown in Table 6. The results for the first shift are presented in Fig. 6.

In this case, two nurses can respond to patient needs at a service level of 82.9 %, three nurses assure a service level of 99.7 %, while the service level is the highest with four nurses. It is also interesting to see that in 0.8 % of simulation repetitions, four nurses were needed and loaded between 60 and 70 % (Fig. 6b).

Table 6 Patient data for 1 day of the case study—"unloaded unit"

Patient p	Barthel scale b_p	Day of the pathway. h_p	Time of patient entry	Presence of patient $w_{s,p}$		
				1	2	3
1	15	1	15/01/2015 16:00	1	1	1
2	20	2		1	1	1
3	12	2		1	1	1
4	18	3		1	1	1
5	14	4		1	1	1
6	13	8		1	1	1

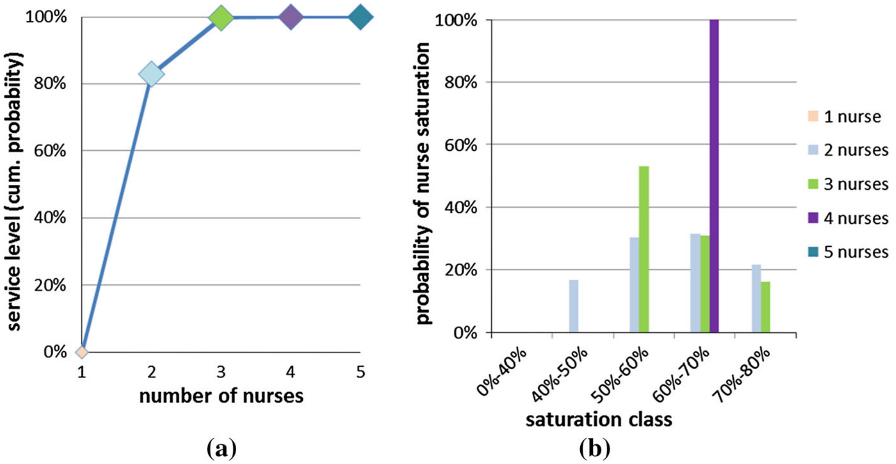

Fig. 6 "Unloaded unit" scenario results: **a** patient service level depending on the number of nurses required n_s per shift s; **b** nurse saturation probability in shift 1

Finally, comparing the two scenarios, it can be stated that in the case of an unloaded unit, two nurses are acceptable, whereas in the loaded unit scenario, three or four nurses are a better choice from the service level point of view. From this comparison, the variability of the loads over the days and over the shifts and the usefulness of the tools in decisions making can be recognized.

5 Conclusions

This study addresses the proposal for a day-to-day nurse requirement planning (NRP) method based on the simulation of patient flow, which takes into account real patient needs considering both the variability of clinical pathways and the duration of care activities. The method incorporates in summary form the best features of both timed-task/activity methods (single tasks of the clinical pathway and the related probability of occurrence are taken into account) and acuity-quality methods (patient dependence on nurses is used to estimate single task duration). Given the probabilistic dimension of the results, both in terms of service level and personnel saturation, it can provide a valid support mechanism for management in nurse rescheduling decisions. The approach is based on patient clinical data collected in Electronic Medical Records and hospital information systems and can work without any human intervention. The validation and verification of the proposal have been undertaken in a stroke unit of a medium-sized acute university hospital, the data for which are reported.

Essentially, a parallel between the "clinical care pathway" and the "manufacturing production process" has been assumed and demonstrated practically in the planning of one type of resource (nursing) considering the stochastic nature of the tasks to be performed and their duration.

The assumption of a flexible workforce has been adopted as it is currently practised in many hospitals. In the case of floating nurses—within a hospital's buffer capacity and dynamically allocated to different medical units—wages are likely to be more expensive due to their higher specialization and cross-training (Kortbeek et al. 2015). This aspect needs to be investigated in depth in further studies, comparing the savings from assuring a high service level (also rescheduling fewer nurses than planned) to the costs of paying for more qualified nurses.

For the sake of simplicity, the model has been formalized by considering the activities of one clinical pathway carried out in one medical unit. However, it can easily be extended to many pathways, medical units and resources. The emerging trend for the use of statistical or empirical modelling methods in estimating the routing of patient flows will make available a sizable amount of data, allowing the successful implementation of the proposed approach in larger departments dealing with many different pathways.

Finally, although the formalization has employed a short-term planning horizon, the method can also be applied, with some slight modifications, to the medium and long term, supporting staffing decisions based on statistics relating to patients' requirements. From this perspective, the method can be considered one "wheel" of a Hospital Resource Planning system (a kind of Enterprise Resource Planning making it possible to plan, manage and control hospital operations), as has often been conceptualized but never fully realized by many authors and practitioners.

References

Adeyemi S, Chaussalet TJ, Xie H, Asaduzaman M (2010) Random effects models for operational patient pathways. J Appl Stat 37:691–701

Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH (2002) Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. JAMA 288:1987–1993

Bard JF, Purnomo HW (2004) Real-time scheduling for nurses in response to demand fluctuations and personnel shortages. In: Proceedings of the 5th international conference on the practice and theory of automated timetabling, pp 67–87

Bayer S, Petsoulas C, Cox B, Honeyman A, Barlow J (2010) Facilitating stroke care planning through simulation modelling. Health Inform J 16:129–143

Bhattacharjee P, Ray PK (2014) Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: a review and reflections. Comput Ind Eng 78:299–312

Brailsford S, Vissers J (2011) OR in healthcare: a European perspective. Eur J Oper Res 212:223–234

Brander P, Norton V (1991) Nursing workload measures in New Zealand public and obstetric hospitals. Information Centre Ministry of Health, Wellington

Burke EK, Curtois T (2014) New approaches to nurse rostering benchmark instances. Eur J Oper Res 237:71–81

Cardoen B, Demeulemeester E (2008) Capacity of clinical pathways—a strategic multi-level evaluation tool. J Med Syst 32:443–452

Chemweno P, Thijs V, Pintelon L, Van Horenbeek A (2014) Discrete event simulation case study: diagnostic path for stroke patients in a stroke unit. Simul Model Pract Theory 48:45–57

Clark R, Walker H (2011) Nurse rescheduling with shift preferences and minimal disruption. J Appl Oper Res 3:148–162

Elf M, Putilova M, Ohrn K, von Koch L (2009) Development of a model of a stroke care process. Online J Nurs Inf 13(3):1–34

Ernst T, Jiang H, Krishnamoorthy M, Sier D (2004) Staff scheduling and rostering: a review of applications, methods and models. Eur J Oper Res 153:3–27

Ghosh B, Cruz G (2005) Nurse requirement planning: a computer-based model. J Nurs Manag 13:363–371

Gillespie J, McClean S, Scotney B, Garg L, Barton M, Fullerton K (2011) Costing hospital resources for stroke patients using phase-type models. Health Care Manag Sci 14:279–291

Gnanlet A, Gilland WG (2009) Sequential and simultaneous decision making for optimizing health care resource flexibilities. Decis Sci 40:295–326. doi:10.1111/j.1540-5915.2009.00231.x

Grieve R, Hutton J, Bhalla A, Rastenytë D, Ryglewicz D, Sarti C, Lamassa M, Giroud M, Dundas R, Wolfe CDA (2001) A comparison of the costs and survival of hospital-admitted stroke patients across Europe. Stroke 32:1684–1691

Guida R, Iannone R, Miranda S, Riemma S, Sarno D (2012) From patients' needs to hospital pharmacy management—an holistic approach to the process modelling. In: Proceedings of the international workshop on innovative simulation for healthcare 2012, part of the international multidisciplinary modeling and simulation multiconference 2012, I3M 2012, pp 56–62

Harper PR (2002) A framework for operational modelling of hospital resources. Health Care Manag Sci 5:165–173

Hurst K (2003) Selecting and applying methods for estimating the size and mix of nursing teams: a systematic review of the literature commissioned by the Department of Health. Nuffield Institute for Health

Hurst K (2005) Relationships between patient dependency, nursing workload and quality. Int J Nurs Stud 42:75–84

Iannone R, Miranda S, Riemma S, Sarno D (2011) Proposal of a conceptual framework to optimise drug management in healthcare systems. In: Proceedings of the IADIS international conference e-Health 2011, part of the IADIS multi conference on computer science and information systems 2011, MCCSIS 2011, pp 215–220

Iannone R, Lambiase A, Miranda S, Riemma S, Sarno D (2013) Modelling hospital materials management processes. Int J Eng Bus Manag 5:1–12

Iannone R, Miranda S, Riemma S, Sarno D (2014) Pulling drugs along the supply chain: centralization of hospitals' inventory. Int J Eng Bus Manag 6:21. doi:10.5772/58939

Iannone R, Lambiase A, Miranda S, Riemma S, Sarno D (2015) Cost savings in hospital materials management: look-back versus look-ahead inventory policies. Int J Serv Oper Manag 22(1):60–85

Jelinek RC, Kavois JA (1992) Nurse staffing and scheduling: past solutions and future directions. J Soc Health Syst 3:75–82

Kaliszer JSM (1976) A survey of patient-nurse dependency. Ir Med J 69:488–492

Kaplan RS, Anderson SR (2004) Time-driven activity-based costing. Harvard Bus Rev 82(11):131–138

Kortbeek N, Braaksma A, Burger CAJ, Bakker PJM, Boucherie RJ (2015) Flexible nurse staffing based on hourly bed census predictions. Int J Prod Econ 161:167–180. doi:10.1016/j.ijpe.2014.12.007

Liang B, Turkcan A (2015) Acuity-based nurse assignment and patient scheduling in oncology clinics. Health Care Manag Sci. doi:10.1007/s10729-014-9313-z

Massachusetts Hospital Association (2010) Hospital costs in context: a transparent view of the cost of care

Mathauer I, Wittenbecher F (2013) Hospital payment systems based on diagnosis-related groups: experiences in low- and middle-income countries. Bull World Health Organ 91:746–756

Molema JJW, Groothuis S, Baars IJ, Kleinschiphorst M, Leers EGE, Hasman A, van Merode GG (2007) Healthcare system design and part time working doctors. Health Care Manag Sci 10:365–371

Naidu KD, Sullivan KM, Wang PP, Yang Y (2000) Managing personnel through staff scheduling algorithms. Proc Jt Conf Inf Sci 5:829–835

Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K (2002) Nurse-staffing levels and the quality of care in hospitals. N Engl J Med 346:1715–1722

Punnakitikashem P, Rosenberger JM, Behan DB (2008) Stochastic programming for nurse assignment. Comput Optim Appl 40:321–349

Rudd AG, Jenkinson D, Grant RL, Hoffman A (2009) Staffing levels and patient dependence in English stroke units. Clin Med 9:110–115

Schroeder RE, Rhodes AM, Shields RE (1984) Nurse acuity systems: CASH vs. GRASP (a determination of nurse staff requirements). Nurs Forum 21:72–77

Siferd SP, Benton WC (1994) A decision mode for shift scheduling of nurses. Eur J Oper Res 74:519–527

Wittenauer R, Smit L (2013) Priority medicines for Europe and the world—2013 update. Background paper 6.6—Ischaemic and Haemorrhagic Stroke. WHO

Zander K (2002) Integrated care pathways: eleven international trends. J Integr Pathw 6:101–107

Debora Sarno received the M.Sc. degree in management engineering and the Ph.D. degree in engineering and economics of innovation from the University of Salerno, Salerno, Italy, in 2010 and 2014, respectively. She is a research assistant in Healthcare Operations Management and a Project Management Professional with Project Management Institute (PMI), USA.

Maria Elena Nenni is Assistant Professor at University of Naples Federico II - Department of Industrial Engineering. She received a master degree with honour in Industrial Engineering from the University of Rome Tor Vergata and a Ph.D. from the University of Naples. She also has been research fellow at CNR – the Italian National Centre for Research Activity. Since 2004 she has been teaching Industrial Plant, Operations Management and Project Management at the University of Naples and Operations Management at the LUISS Guido Carli University for the graduation classes. She is also lecturer in several master and management training courses at LUISS Business School (LBS). She is member of the LBS Knowledge Centre and she is Scientific Coordinator for the topics of Operations and Project Management.