Health Services Research

24

Thomas Schäfer, Christian A. Gericke, and Reinhard Busse

Contents

T. Schäfer (✉)
Department of Economics and Information Technology, University of Applied Sciences Gelsenkirchen, Bocholt, Germany

C.A. Gericke
The Wesley Research Institute and University of Queensland, School of Population Health, Brisbane, Australia

R. Busse
Department of Health Care Management, Berlin University of Technology, Berlin, Germany

W. Ahrens, I. Pigeot (eds.) *Handbook of Epidemiology, 2nd edition*,
DOI 10.1007/978-0-387-09834-0_38,

24.1 Introduction

After a brief introduction into the general field of health services research, a large section deals with the specific issues arising when epidemiological or statistical methods are used to study health services. This is followed by sections describing the main fields of investigation which are usually thought of as pertaining to the wider realm of health services research. These are studies of demand, need, utilization, and access to health services which have the interface between the patient and health services in common. The next section describes the importance of financial resources, structure, and organization for the delivery of effective and efficient health care. This is followed by a description of the processes and outcomes of health care, including concepts such as effectiveness and appropriateness of care and their use, for example, in physician profiling or in hospital rankings. In the section on outcomes, special emphasis is put on health status measurement and the evaluation of health systems in international comparisons. Important health economic concepts, such as cost-effectiveness and efficiency, are covered in various sections. This chapter concludes with describing common pitfalls and caveats in interpreting health services research.

24.1.1 Health Services Research Defined

Health services research (HSR) attempts to answer questions about the best medical treatment or preventive course of action, the quality of care provided by a hospital or a physician, the efficient delivery of services to all populations, and their costs. The Institute of Medicine (1994) defines HSR as "A multi-disciplinary field of inquiry, both basic and applied, that examines access to, and the use, costs, quality, delivery, organization, financing, and outcomes of health-care services to produce new knowledge about the structure, processes, and effects of health services for individuals and populations." The three basic dimensions of care studied are (1) the process of deciding what care to provide, (2) the process of providing care in the best possible manner, and (3) the outcomes that result from care (Scott and Campbell 2002). Many HSR projects study aspects of care that span all three dimensions under the rubric "quality of care" (Brook and Lohr 1985). As Scott and Campbell (2002) pointed out, this frequently used but rarely defined phrase encompasses notions of effectiveness, efficiency, safety, access, and consumer satisfaction and is thus not a very precise title for scientific investigation (Scott and Campbell 2002). HSR challenges the dominant biomedical model in which disease occurs, leading to illness, which is then treated (Black 1997). In contrast to the clinical view focusing on individual patients, it adopts a population perspective and considers other determinants of the use of health care (Black 1997), such as socioeconomic status, local availability or acceptability of health services. HSR thus often challenges medical claims about the value of specific interventions.

HSR cannot be defined as a methodological discipline. It draws upon and uses multiple methodologies and is multidisciplinary in nature. The majority of quantitative research in the field is done using epidemiological methods, and epidemiologists increasingly work in this field of research.

This multidisciplinary approach is seen by many authors as characteristic of this field of investigation, which is reflected in Last's definition of HSR as "The integration of epidemiological, sociological, economic, and other analytical sciences in the study of health services. HSR is usually concerned with relationships between need, demand, supply, use, and outcome of health services. The aim is evaluation, particularly in terms of structure, process, output, and outcome" (Last 2001).

The ultimate goal of HSR, however, is to provide unbiased, scientific evidence to influence health services policy at all levels so as to improve the health of the public (Black 1997).

24.1.2 The Input-Output Model of Health Care

Different models have been proposed for the study of health services. These include operational models, for example, the patient-flow model or the social sciences model. The patient-flow model starts with the assumption of a healthy population, where a patient's way through the different health-care institutions is followed once a disease manifests itself (Bennett 1978). The social sciences model attempts to consider the main social and political influences, causal relationships, and environmental conditions on the process of service delivery in a health-care system. Social experiences, values, priorities, importance of societal resources and structures are the focus of the analysis (Weinermann 1971). A causal or epidemiological model is also possible, which analyses care along known or supposed hypothetical causal biosocial links (de Miguel 1971). The drawback of such a model is its complexity.

For many analyses, simpler models are more adequate. We prefer a model adapted from engineering sciences in which some components of the other models have been integrated – the input-output model (Fig. 24.1) – which takes structure, processes, outputs, and outcomes into account (Schwartz and Busse 2003).

In this model, statistical data can be structured in an easy and transparent manner. Political debates on health services also often follow this structure.

The input of the health-care system is divided into (Schwartz and Busse 2003):

- Patient-side input, that is, the health status of the population as well as its access to care
- Resource oriented input, that is, the input in terms of financial and non-financial resources, such as human resources and infrastructure, as well as organizational structures, responsibilities, and interdependencies between actors and organizations

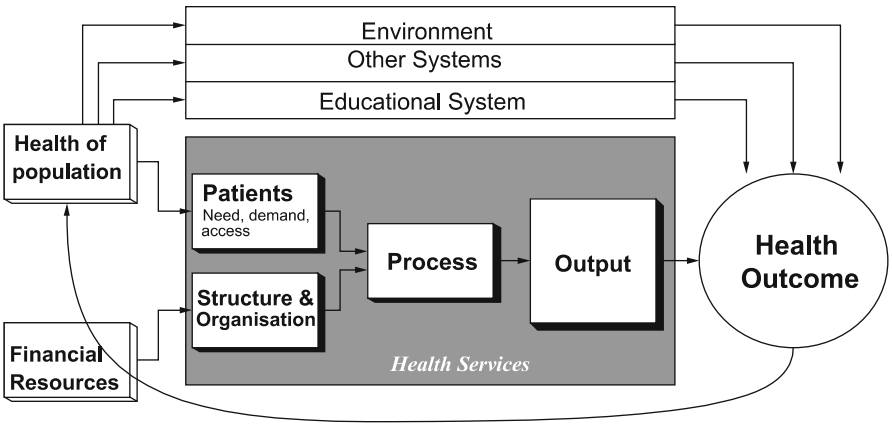

Fig. 24.1 The input-output model (Busse and Wismar 2002)

Throughput forms the center of the model – encompassing all processes of care in a health-care system.

The output of the health-care system is divided into two sequential elements (Schwartz and Busse 2003):

- Direct results of the processes, that is, output measures in the classical sense, also termed intermediary outcomes, for example, the number of cardiac catheterizations performed
- Outcomes in terms of changes in health status, which are often only measurable in the long-term, for example, the mortality avoided by a specific intervention

Common to all models deployed is the problem of causal inference. Although some problems are also encountered by epidemiological research, like the establishment of precedence in time in case-control studies or in historical cohorts (Hill 1965), these problems are much more important in health services research and thus make the latter more prone to biased or confounded results. The main problem is the complexity of the system with multiple interdependencies which result in the dilemma of "before the intervention is after the intervention." A good example is the evaluation of health-care reforms, which often come in a piece-meal fashion and which are only half-way executed before the next reform measures start. Assigning observed changes in an evaluation study to one particular reform package then becomes difficult, and as with all uncontrolled before-and-after studies, the results of such studies have to be interpreted with great caution (Grimshaw et al. 2001). However, in particular, for the evaluation of health reforms, such before-and-after studies are often the only possible research method, as conducting controlled studies is often not feasible.

24.1.3 Level of Analysis

A complementary approach to the one described is the analysis of the level at which processes of care take place (Schwartz and Busse 2003):

- The macro level – consisting of the health system as a whole and national health policy
- The meso level – research focusing on interorganizational structures and processes, for example, between health-care payers and providers, or the relationships between providers in a specific region
- The micro level – analysis of individual care services and technologies

Aday et al. (1998) attempted to match research methods to the level of analysis, illustrated in Table 24.1. In their model, the macro level refers to a population perspective on the determinants of the health of communities as a whole ("health of population" in the model), and the micro level represents a clinical perspective on the factors that contribute to the health of individuals at the system, institution, or patient level (Aday et al. 1998). Their intermediary system level encompasses both the macro and meso level in our model. It refers to the resources (money, people, physical infrastructure, and technology) and the organizational configurations used to transform these resources into health-care services either for the country as a whole (macro level) or within a specific region (meso level).

Table 24.1 Levels of analysis in health services research (Adapted from Aday et al. (1998))

	Level of analysis			
Data sources	Population	System	Institution	Patient
Census	x			
Public health surveillance systems	x			
Vital statistics	x			
Surveys				
Population	x			
Organizations		x	x	
Providers		x	x	
Patients		x	x	x
Insurance records/ administrative data				
Enrollment	x	x	x	x
Encounters		x	x	x
Claims		x	x	x
Medical records		x	x	x
Qualitative studies				
Participant observation	x	x	x	x
Case studies		x	x	x
Focus groups	x		x	x
Ethnographic interviews	x			x

24.2 Methodological Considerations

Generally all types of data can be analyzed for the purposes of HSR. We find experimental data from randomized controlled trials as well as observational data from case-control or cohort studies, registers, or surveys. But many analyses in HSR make use of data from large administrative databases that are abstracted from medical or hospital discharge records, prescriptions, and bills of payments for delivered health services.

24.2.1 Study Designs

24.2.1.1 Randomized Controlled Trials

When alternative approaches to the delivery of health care have to be evaluated, a randomized controlled trial (RCT) is considered the gold standard, that is, the most rigorous method available. RCTs are performed in order to avoid bias and confounding. It is the effect of randomization that provides equality of study and control group in all relevant characteristics except the intervention being tested.

Regardless of this advantage, one has to keep in mind some critical issues when considering the RCT methodology. First, the recruitment of participants (or other experimental units like communities or schools, etc.), who meet all eligibility criteria, may be difficult and expensive. Furthermore, a randomized assignment of treatments to patients may not be feasible by ethical reasons (for instance, if you want to compare a treatment that is widely believed to be efficacious with "no treatment" or with a placebo). The study population is frequently not representative of the target population. Thus, it is true that an RCT has a high level of "internal" validity, as study group and control group are really comparable, but this tends to be connected with a low level of "external" validity which is an important consideration in HSR as it aims to examine effects under actual conditions and not under trial conditions. The results of an RCT refer solely to efficacy (a treatment is called efficacious if the desired effect is obtained under optimal conditions) but not necessarily to effectiveness (a treatment is called effective if the desired effect is obtained under everyday conditions).

Accordingly, the role of RCTs in HSR is more limited than in other areas of health research, and RCTs have only been carried out in certain areas of HSR. For example, the efficacy of cholesterol-lowering treatment in the prevention of coronary heart disease in men with high cholesterol was demonstrated by a multicenter, randomized, double-blind clinical trial, the *Lipid Research Clinics Coronary Primary Prevention Trial* (Lipid Research Clinics Program 1984). As Kelsey et al. (1998) report, the most expensive research study ever sponsored by the US National Institutes of Health – the Women's Health Initiative (Buring and Hennekens 1992) – consists of a series of RCTs to test the hypotheses, whether a low-fat dietary pattern protects against breast cancer and colon cancer, whether hormone replacement therapy reduces risk for coronary heart disease, and whether calcium and vitamin D

supplementation protects against hip fractures. Even in the context of evaluating organizational change, RCTs had been carried out. The so-called *Health Insurance Experiment* (Newhouse 1974) was designed as an RCT to evaluate the effect of different levels of cost sharing in health insurance on utilization, expenditures, and health status (for more details, see Sect. 24.3.2). But regardless of such examples, the majority of RCTs are designed to evaluate a new (drug) therapy and are performed in clinical settings (randomized *clinical* trials).

24.2.1.2 Observational Studies

An overwhelming part of HSR is based on observational studies. In a pure epidemiological setting, case-control and cohort studies are used to estimate and evaluate the association between a specific exposure and a specific disease. In addition, exposure and outcome are frequently mapped into binary variables. In HSR, exposures and outcomes have a higher degree of variety than in chronic disease epidemiology.

Typical exposures are

- Conditions that may lead to inequalities in access to care, for example, low-income status, rural area of residence
- Health states that define certain needs for care, for example, mental illness
- Medical interventions, for example, stent implants versus bypass surgery to prevent heart attacks
- Different health-care delivery systems, for example, Health Maintenance Organizations (HMO) versus capitated Preferred Provider Organizations (PPO) versus a traditional indemnity plan with fee for service (FFS) payments
- Programs that aim to improve the quality of care, for example, disease management programs
- Programs to contain the costs of care, for example, drug formularies

Typical study outcomes are

- Access to care, for example, preventive services (vaccination), therapeutics
- Health status, for example, incidence of prespecified (tracer) diagnoses
- Life years gained, that is, reduction of mortality
- Patient reported outcomes, for example, health-related quality of life
- Quality of care scores, for example, measures of the Health Plan Employer Data and Information Set (HEDIS)
- Appropriateness of care measures, for example, an appropriateness evaluation protocol (AEP)
- Cost of care, for example, increases (additional costs or losses) or decreases (savings)

The general limitations of observational studies are dealt with among others in chapters ▸Cohort Studies, ▸Case-Control Studies, ▸Modern Epidemiological Study Designs, ▸Confounding and Interaction and ▸Design and Planning of Epidemiological Studies of this handbook. The absence of randomization gives reason for special concerns, that is, a special type of selection bias, when the goal of the study is to evaluate interventions. Persons who choose a particular intervention – or are advised by a physician to undergo it – are often on a different level of risk

for the outcome of interest compared with persons who are not assigned to or did not choose to use this intervention (Selby 1994). Particularly in the case of an open intervention program, persons who pay attention to their health may be more likely to participate and comply with a recommendation (e.g., to undergo a screening examination) than persons who do not (Kelsey et al. 1998).

Case-Control Studies The methodology of case-control studies is treated in great detail in chapter ▸Case-Control Studies of this handbook. Because of its obvious merits (a case-control study can be carried out at relatively low cost and comparably quickly), it is used with increasing frequency in HSR, especially in order to asses the adverse effects of drugs and other therapies and to evaluate the efficacy of preventive interventions (Kelsey et al. 1998).

Examples for the application of the case-control approach in HSR are numerous. This includes evaluations of vaccine efficacy and vaccination effectiveness, assessments of medical therapies, of screening programs for cervical, breast, and colon cancers, and of a number of programmatic activities in the community (Armenian 1998).

Cohort Studies Primary data collection for classical epidemiological cohort studies (cf. chapter ▸Cohort Studies of this handbook) is relatively rare in HSR compared to chronic disease epidemiology. One reason might be that health systems change fast in small scales, that is, in trends in coding diagnoses, and in large scales as, for example, completely new reimbursement structures like diagnosis-related groups (DRGs) so that information from long-lasting follow-up studies may often be outdated and not worth the expense.

The majority of cohort studies in HSR is based on administrative data collected for purposes other than research and is focused on the outcomes of a medical treatment or a preventive intervention. The outcomes vary and may include mortality, morbidity, functional status, quality of life, costs, and satisfaction with care. The studies frequently use historical cohorts. For example, the investigation of short-term (30-day) and long-term (5-year) mortality in a cohort of members of a large HMO with hip fractures was a historical study that used computer-stored hospital discharge data linked with computer-stored data from death certificates (Petitti and Sidney 1989). A different type of cohort analysis in HSR focuses on the description of changes in symptoms, functional status, or quality of life in patients who undergo a treatment or are the subject of a preventive intervention (Petitti 1998a).

Cross-Sectional Studies If the goal of data analysis is related to health planning or the assessment of needs for services, prevalence rates are often more useful than incidence rates. Cross-sectional studies therefore represent an important tool for health planning and evaluation. In outcome research, the common methodological approach of *variance in practice* (e.g., to assess the quality of medical care or the outcome of the health system of a county) is tightly connected to a cross-sectional design, mostly based on administrative data, with organizations (e.g., hospitals), providers (e.g., surgeons), counties, states, or even countries as the units of analysis.

Cross-sectional studies are also used to establish research priorities based on consideration of the burden of disease (Kelsey et al. 1998). In a study on the prevalence of chronic gynecologic conditions among US women of reproductive age, for example, it was found that the most common conditions were menstrual disorders, adnexal conditions, and uterine fibroids. The results stressed the need for more effective treatments for these disorders and moreover, suggested that more research on their etiology would be highly desirable (Kjerulff et al. 1996).

Cross-sectional studies are of course less useful to examine hypotheses on causal effects mainly because of the lack of knowledge on the temporal sequence of hypothetical causes and potential effects but also because cross-sectional studies include both new and old cases. This results in a case group which has more than its fair share of individuals with disease of long duration because those who die or recover quickly will be underrepresented (Kelsey et al. 1998).

24.2.2 Complex Models for Data Analysis

In several health systems, available claims data are characterized by a longitudinal structure with long strings of repeated measures of health services for individual patients. Such data structures demand analytical designs. To make full use of them, complex longitudinal data analysis techniques must be applied that can handle time-varying exposures, repeated outcomes, and intra-person correlations.

The lack of detailed information on the severity of disease in claims data sometimes is a reason to use case-based study designs, as for instance, case-crossover studies to allow cases to be their own controls (cf. chapter ▸Intervention Trials of this handbook).

Another complicating factor is that observations in health-care delivery systems are often not independent. For many observational studies, the level of observation is a patient (characterized by a vector of patient attributes). A cluster of patients will be seen by the same physician (characterized by a vector of physician attributes) and will therefore experience similar treatment patterns so that their outcomes cannot be expected to be completely independent. Physicians often practice in groups sharing similar practice styles. These groups may practice in a larger health-care delivery system that imposes constraints to treatment choices, for example, drug formularies or payment by capitation (a lump sum per patient), which will make practice styles of groups within a health plan more similar than groups outside the plan. This clustering of observations on multiple levels has led to the adoption of multilevel regression models as standard tools of HSR.

24.2.3 Data Sources

Primary data collection in a randomized controlled trial, a case-control or a cohort study, is certainly an important, although unusual, data source of HSR. Primary data, when used in HSR, are more frequently collected from the general population (or subgroups) by questionnaire. The majority of data that are analyzed in HSR stem from large administrative data bases, as pointed out before.

24.2.3.1 Surveys

Survey research is frequent in HSR. For a detailed description of survey methods, see chapter ▸Epidemiological Field Work in Population-Based Studies of this handbook. On the one hand, it can be used to provide snapshots of the current state of a health-care delivery system. On the other hand, survey subjects often become re-interviewed in regular intervals to form a longitudinal data structure or a panel. Further possibilities to classify surveys are given by the

- Unit of observation (patients, patient-provider contacts or providers)
- Target population (total population or subgroups)
- Type of data collected (interview data, data of medical examination, or both)
- Access to information (personal interview, mail survey, or telephone interview)

A well-known German survey – the *EVaS-Study* (*Study among office-based ambulatory care physicians in the Federal Republic of Germany*, Schwartz and Schach 1989) – was a cross-sectional survey with patient-physician (or patient-office) contacts as units of observation. The concept followed the US *National Ambulatory Medical Care Survey* (*NAMCS*, Tenney et al. 1974) to some extent. The target population was defined by a number of selected regions in Germany, a fixed study period, and the exclusion of a few medical specialties concerning the involved physicians. The data were collected by mail using an induction interview questionnaire, a reporting form, and a final questionnaire. The final data record covers data of the patient as well as data provided by the physician's office (including, e.g., the diagnosis corresponding to the patient's major reason for encounter, the assessment of the severity of the problem, the services delivered, and the duration of the encounter).

The *German National Health Interview and Examination Survey (GHS)*, however, (carried out from October 1997 to March 1999) targeted the general population aged between 18 and 79 years (Bellach et al. 1998).[1] The units of observation had been residents who were interviewed and medically examined. The data are available for research as a public use file. One of the results of this survey, for example, concerned the utilization of medical services available in Germany under statutory sickness fund facilities. About 90% of all Germans had seen their doctor at least once a year. Half of the population had consulted a doctor during the past 4 weeks and, on average, a medical practitioner was consulted 11 times a year (Bergmann and Kamtsiuris 1999). The Robert Koch Institute continued the GHS with the *German Health Interview and Examination Survey for Adults*. The recruitment period for this survey is November 2008 until November 2011 involving a total of 180 cities and municipalities all over Germany (Robert Koch Institute 2008a).

[1] This survey had three predecessors in the years 1984–1986, 1987–1989, and 1990–1991 and was supplemented by the *German National Health Interview and Examination Survey for Children and Adolescents (KiGGS),* carried out from May 2003 to May 2006. The target population were children and adolescents aged between 0 and 17 and living in Germany (Kurth 2007). The Robert Koch Institute continued the KiGGS study by carrying out telephone-based health interviews (Robert Koch Institute 2010).

Another frequently cited German survey is based on a representative, regionally stratified sample of 0.4% of all prescription forms, which are completed by office-based physicians for members of statutory sickness funds. This survey is supported in cooperation by the federal associations of the office-based physicians, the statutory sickness funds, and the free-standing pharmacies. It is carried out each year. The annually published results include an analysis of the sales increase with respect to its components referring to prices, volumes, and structural composition (Schwabe and Paffrath 2010).

The US *Medicare Current Beneficiary Survey (MCBS)* is an example of a survey that is designed as a panel. MCBS began in 1991 as a continuous panel in order to provide a more complete picture of the use of health services, expenditures, and sources of payment for the Medicare population. It is an ongoing computer-assisted personal survey of Medicare beneficiaries residing in the United States and Puerto Rico. Each person is interviewed three times per year over 4 years (regardless of whether he or she resides in the community or a long-term care facility), following a 4-year rotating panel design. The MCBS thus contains four overlapping panels of Medicare beneficiaries. Each year, one panel is dropped from the survey, and a new one is added. This design produces three calendar years of medical utilization data for each sample person. The data are collected over a 4-year period in which sample persons are interviewed 12 times. The first interview collects baseline information on the beneficiary. The next 11 interviews are used to collect three complete years of utilization data. Included are medical expenditure data as well as detailed data on health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment (Adler 1994). The data are used to produce calendar year public use files on access to care, and cost and use. The nationwide MCBS data are released – as usual for public use files – only under a data use agreement. In addition, requests for regional or supplementing data must include a study protocol with specific justification for the additional data required, along with an identifiable data use agreement (see http://cms.gov/mcbs).

Another excellent population-based US panel, created by the (former) Agency for Health Care Policy and Research and the National Center for Health Statistics, is the *Medical Expenditures Panel Survey (MEPS)* which collects data from several sources to provide a complete picture of the health status and health-care utilization of a random sample of citizens (Cohen 1997).

In addition to other sampling methods, computer-assisted telephone interviews have become more frequently used in HSR. This method has comparatively low costs and guarantees an approximate full coverage of the general resident population in developed countries which have high rates of telephone access. Even unlisted households can be covered by means of random digit dialing.[2] Data are checked for correctness, completeness, and plausibility and stored continuously in the

[2]But the increased use of cellular phones poses a problem, and there is need for research to broaden the approach beyond the restrictions of the conventional telephone network.

course of the interview. Separate steps for data input and examination are not necessary. Germany, for example, started the first *National Telephone Health Survey* in September 2002. About 8,000 German speaking residents aged 18 years and older had been questioned on diseases, health-related behavior, and utilization of the health-care system (Ziese et al. 2003). This survey was supplemented by a regional one in Bavaria (Meyer et al. 2002). The recent telephone survey carried out by the Robert Koch Institute, the *German Health Update*, began in July 2008 and ended in April 2009. Twenty-five thousand people aged 18 and older were selected for an interview. The German Health Update was supplemented by regional surveys in Brandenburg and Saarland (Robert Koch Institute 2008b).

24.2.3.2 Official Statistics

Official mortality and other health or demographic statistics, especially vital statistics (births, marriages, deaths, etc.), have been extensively used in HSR. An early well-known and frequently cited example in the context of equity research is the study on differential mortality in the United States (Kitagawa and Hauser 1973). The comparison of mortality rates and the proportions of money from the national accounts that are spent on health is a popular starting point for health economists in analyzing the efficiency of a particular health system. But official statistics – mostly based on law – resemble routine registries and share their limitations (cf. chapter ▸Use of Health Registers of this handbook and Sörensen 2001). The validity of mortality statistics in particular is strongly dependent on the rate of autopsies in a country. In a German survey of institutes of pathology within universities and in community hospitals in the year 2000, a median value of 23.3% and 13.3% of autopsies among hospital deaths was found, respectively. This was considered clearly below the recommended value of 30% (Schwarze and Pawlitschko 2003).

24.2.3.3 Administrative Databases

Administrative data are abstracted from medical or hospital discharge records, prescriptions, and bills of payments. Thus they have several advantages. They are routinely collected data representing the reality of health-care delivery. They need no additional time and money to gain access to large patient populations over long periods of time with repeated recordings of most health-care encounters of each subject. But the advantage of quick and easy access to large and representative populations is counterbalanced by data that may be incomplete and suffer from voluntary and involuntary miscoding. Although the quality of these data appears to improve over time, it has to be kept in mind that the primary reason for creating them was to document medical diagnoses and interventions obtained from medical records and manage the flow of payments for delivered health services obtained from claims data.

Since the advantages, particularly of electronic claims data, are so obvious, researchers try to better understand the consequences of the data limitations and develop analytical methods to adjust for them. Say, for example, the approach of Newhouse and McClellan (1998) used to overcome the typical selection problem,

they were confronted with in analyzing the data of catheterization of patients with acute myocardial infarction. As data limitations are unique to each administrative database, a very good understanding of how data were generated is crucial for interpreting analytical findings.

Examples of administrative databases in the USA include the national Medicare and Medicaid databases as well as claims files for privately insured patients or members of a particular health facility. Data from the Medicare program, run by the *Centers for Medicare & Medicaid Services (CMS)*, are confidentiality-protected, longitudinally linked, person-level records that track virtually all elderly US citizens from their 65th birthday onward until death, through geographical moves and changes in providers.[3] The data sets include the types and amounts of health services used (e.g., hospitalizations, office visits, home health care, surgeries, and diagnostic tests), the medical problems being treated (diagnoses), provider characteristics (site of service and physician training), and charges. Information on long-term care services and outpatient prescription drugs, not covered by CMS, is not included (Diehr et al. 1999).

In Germany, administrative databases which do not find their ways into the official statistics or any kind of survey are scattered across the statutory sickness funds or other agencies of social security. Due to comparably strict data protection rules, record linkage is not a common practice in Germany.

From 2004 onward however, the associations of sickness fund physicians have been obliged by law to transfer beneficiary-related billing data, including diagnoses, to the statutory sickness funds.[4] Since then, anonymized/pseudonymized beneficiary-related databases from different research institutes were established, which encompass the data of several insurance companies and are used for purposes of drug safety and health service research (cf. Grobe et al. 2006, 2011; Ihle et al. 2008; Pigeot and Ahrens 2008; Glaeske and Schicktanz 2010; Bitzer et al. 2010; Sauer et al. 2010; Rothgang et al. 2010; Schäfer et al. 2011). In addition, the Federal Insurance Office governs a sample, which is annually drawn from all statutory insured persons, used for the calculations on risk structure equalization.

The abundance of information in claims databases in various states is often overwhelming. Many redundant measures are recorded, and researchers must identify the underlying variables that represent the concepts they want to evaluate. Since researchers on the other side, hate to discard already recorded information, *data reduction techniques* including comorbidity scores or propensity scores are increasingly applied to condense data while preserving information. For a detailed discussion of pharmacoepidemiological databases, we refer to chapter ▸Screening of this handbook.

[3] About 10% of Medicare enrollees are younger, disabled persons, who are tracked from their time of certification.

[4] The hospitals and the pharmacies had started to transfer their beneficiary-related data to the sickness funds long before.

24.2.4 Measurement Error (Misclassification)

In HSR, measurement errors (non-differential and differential as well; cf. chapter ▸Measurement Error of this handbook) for all variables of interest are considered to be higher than in a traditional epidemiological setting for several reasons:

- Some data that are collected and stored but are not directly used for reimbursement or other administrative purposes (e.g., job or social status of an enrolled person) are likely to be not up-to-date.
- Diagnostic information on claims is documented to justify reimbursement; a bill for tests to rule out cancer, for example, may contain a diagnostic code for "cancer" even if the tests were negative.
- The information on clinical conditions in administrative data is in the form of diagnoses coded using the International Classification of Diseases (ICD) which is revised from time to time. Currently, ICD-10 (the tenth revision) is in use. Several countries (e.g., the USA) prefer ICD-10-CM, a clinical modification of the ICD. Some diseases such as arthritic or psychiatric disorders are difficult to classify because of lack of clearly defined diagnostic criteria. Less serious or vague conditions have a high probability of inconsistent coding. Regional and temporal variations of coding patterns may additionally reduce the reliability of coded diagnostic information.
- Especially the reliability of ambulatory diagnoses is a major concern. An analysis by the Medicare Payment Advisory Commission – MedPAC (1998) – demonstrated the inaccuracy of outpatient diagnosis coding. For the purposes of the study, MedPAC selected beneficiaries whose Medicare Part B claims in 1994 showed a diagnosis of 1 of 11 serious diseases, then checked for claims for the same diagnosis in 1995. As shown in Table 24.2, the likelihood of a claim in 1995 was only about 50–60% for each of the 11 diagnoses (cf. Newhouse et al. 1997). Part B Medicare covers the costs for medical service by general practitioners,

Table 24.2 Persistence in diagnostic coding of those identified in 1994 (Source: Medicare Payment Advisory Commission 1998, p. 17, Note: Excludes those who died in 1994 or 1995)

Diagnosis on 1994 Part B claim	Percent with Part B claim in 1995
Hypertension	59
Coronary artery disease	53
COPD	62
Congestive heart failure	61
Stroke	51
Dementia	59
Rheumatoid arthritis	55
High-cost diabetes	58
Renal failure	56
Quadriplegia/paraplegia	52
Dialysis	59

for a small selection of pharmaceuticals, for ambulatory treatments in hospitals, and other therapeutic and care services that are not covered by Part A Medicare.

- Moreover, administrative data used to reimburse hospitals or physicians are subject to some problems that are not familiar to epidemiologists, called "upcoding," "coding proliferation," and "gaming." Upcoding of diagnoses to more serious conditions is the process of assigning a diagnosis code or codes to a patient that may maximize the provider's reimbursement (e.g., ischemia to myocardial infarction) as it has occurred with some DRG payment systems (Dunn et al. 1996). Coding proliferation means the increase in the coding of all related conditions affecting treatment. Both types of distortion are relevant sources of measurement errors in HSR. Gaming is a serious problem that is dealt with in Sect. 24.5.2.3 because it cannot be subsumed under the term "misclassification" seamlessly.

In summary, the quality of claims data may be adequate for some purposes, but it is important to remember that claims are generated to justify reimbursement rather than to facilitate research.

24.2.5 Sampling Issues

A considerable part of data analyzed in HSR is based on samples from the population of interest (this is called the "target population" or the "population being sampled"). Sampling can help to save time and money. Sampling may also result in an increase of accuracy of measurement since more effort can be spent on this issue if only a manageable number of units of observation is included. Scientifically sound sampling methods are indispensable tools for designing an efficient sample and to provide consistent and unbiased projections from complex samples. Scientific sampling means *probability sampling*, that is, the probabilities of selection must be under control. Non-random samples based on volunteers or on the judgment of the sampler are not covered by this concept and are not recommended for use in HSR.

The sampling procedure is called *simple random sampling* if each of the possible samples of a given size has an equal chance to be selected. It follows that every one of the sampling units in the population has the same chance of being included in the sample. This is occasionally offered as the definition of simple random sampling (e.g., Kelsey et al. 1998) without realizing that there are other sampling procedures (e.g., systematic sampling, see below), which also have this property (Sukhatme and Sukhatme 1970). Simple random sampling is simple in theory but less so in practice because one needs a complete list of the population to draw the sample.[5] In many instances, it may not be the most efficient method of sampling. Therefore – apart from telephone sampling – it is not much used in practice.

[5]Random digit dialing in order to sample for computer-assisted telephone interviews is considered as a way to handle this problem if the existing lists are not complete and the target population can be accessed by conventional telephone network.

Systematic sampling is a common type of sampling based on selecting every *k*th individual from a list or a file after choosing a random number from 1 to k as starting point. It is based on a fixed rule and is not limited to selection from an actual file. Thus, selection of all those born on the (randomly chosen) third day of any month or of everyone whose social security number ends in (the randomly chosen digits) 17, 48, or 76 is similar to systematic sampling procedures yielding approximately 3% samples.[6] Because of these properties, systematic sampling is often simpler to administer under field conditions than simple random sampling. But systematic sampling has a severe handicap: differently from other sample designs, it is impossible to estimate the variance from one single sample. For an unbiased estimate, you need repeated sampling. Several (biased) approximations are used in practice to estimate the variance. One of these consists in treating the systematic sample as if it was a random sample of *n* units (Sukhatme and Sukhatme 1970).

When the population can be divided into *strata* in such a way that each stratum is more homogenous than the population as a whole, one can reduce the sampling error compared to a simple random error. Examples of variables that are often used for stratification are region, age, sex, race, and socioeconomic status. Following a *stratified sampling* design, a separate sample is drawn from each stratum, and the results are then appropriately combined in the analysis.

Cluster sampling has contrasting properties compared to stratified sampling. It is a simple random sample applied to groups of population members (*clusters*) that usually leads to a substantial loss of precision. But because of operational improvements of access to the units to be selected, one often achieves a heavy decrease of collecting costs and thereby an increase in precision per unit of cost. Examples of clusters that could be sampled are hospital wards, villages, schools, families, etc. If clusters are positively correlated within themselves, that is, they have a high positive *intraclass-correlation coefficient (ICC)*, indicating more homogeneity than would result from chance alone, cluster-sampling variance will be larger than simple random sampling variance. This is a situation frequently observed in real life. As ICCs are positive in most cases, simple sampling variance can grossly underestimate the true cluster-sampling variance. The ratio of the latter to the first-mentioned variance is called the *cluster effect*.

In a *multistage sampling design*, stratification and clustering may be combined on several stages of the sampling procedure forming a complex random sample. Stratified sampling, for example, may be used to ensure that schools are represented in the sample according to different socioeconomic areas in a large city, and cluster sampling of classrooms within the selected schools might then be employed for efficiency.

The analysis of data from a complex sample procedure that includes cluster sampling requires a sound knowledge of sampling theory or statistical advice. There

[6]Of course, the choice of the sampling scheme has to be relevant to the population being sampled. A population that is not completely covered by social security would be unsuitable for sampling by means of social security number.

are a series of textbooks on sampling theory (e.g., Hansen et al. 1953; Kish 1965; Stuart 1968; Sukhatme and Sukhatme 1970; Cochran 1968; Levy and Lemesshow 1991), and many handbooks or textbooks on statistics contain at least a chapter on sampling theory (e.g., Kendall and Stuart 1958; Kahn and Sempos 1989; Krishnaiah and Rao 1994; Voß 2003). The use of a special software (e.g., Sudaan) or a special module of one of the common large statistical packages (e.g., SAS, Stata or SPSS) is inevitable. They allow for variance weighting in the statistical procedures to adjust for the specific sampling design. Otherwise, as cluster-sampling variance may be many times larger than the variance calculated by assuming a simple random sample (Abraham 1986), and the analysis can result in severely misleading conclusions about the significance of the study findings.

24.2.6 Confounding and Risk Adjustment

For general principles of control for confounding, see chapter ▸Confounding and Interaction of this handbook. Health services researchers tend to summarize methods to adjust for confounding under the term "risk adjustment." With respect to the large databases analyzed, standardization and multivariate modeling are more frequently used to control for confounding than the traditional approach of stratification.

Any level of comparison can be affected by confounding. This includes the mapping of health-care needs, the evaluation of clinical strategies and programs, studies of the effectiveness of quality improvement initiatives, or the evaluation of cost containment measures. Typical confounders are age, sex, ethnicity, income, smoking, or other risk variables. In outcome studies, confounding is a major concern because of differences in severity of illness and comorbidity.

The most frequent approach to control for confounding in HSR is to include the potential set of confounders as predictors in the regression model (cf. chapter ▸Regression Methods for Epidemiological Analysis of this handbook) to predict the outcome of interest:

- Ordinary least squares (OLS) regression when the outcome has a continuous distribution (ideally a normal distribution) as, for example, the logarithm of costs
- Poisson (or negative binomial) regression when the outcome is described by counts (as, e.g., the number of hospital admissions in a specified year)
- Binomial (logistic) regression when the outcome variable is binary, indicating, for example, the occurrence of disease or death

When a large database is used for the analysis, comorbidity is often taken into account by including a lot of so-called dummy, that is, binary 0/1 variables in the regression model that indicate the presence (or absence, respectively) of each out of a list of classified comorbidities. When using samples of small or moderate size, this approach may not be possible. In this case, a premodeled aggregated index of comorbidity can be included in the analysis (Schneeweiss et al. 2001).

Including a potential confounder variable in the analysis requires its storage in the database. This is crucial whenever "secondhand" data are analyzed. Especially in

a country like Germany where record linkage is not a common practice, one cannot expect to have full control over all relevant confounders, and many socioeconomic characteristics are available only in survey research.

Omission of one or several confounders usually leads to a violation of the assumptions underlying the estimation procedure in the OLS regression model

$$Y_i = \beta_1 x_{i1} + \beta_1 x_{i2} + \ldots\ldots\ldots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \ldots, n,$$

as the k predictors x_{ij}, $j = 1, \ldots, k$, and the random errors ε_i, $i = 1, \ldots, n$, are no longer uncorrelated where n denotes the number of subjects, Y_i the response variables, and β_j the regression coefficients.

In this situation, the introduction of one or more of so-called *instrumental variables* can help to establish a consistent estimation of the interesting effects (that is, β_j). This has been a well-known technique of econometric analysis (IV-technique) for over half a century which is described in almost every econometric textbook (e.g., Greene 2003). But it is very rarely applied to HSR problems because a sound econometrical or statistical background is needed. Instrumental variables should be correlated with the predictor variable as much as possible, and at the same time, they should be (at least asymptotically) uncorrelated with the random errors. The skill of the technique consists in finding such variables which are already in the database or could be added to it at a tolerable cost.

An illuminating example of the usefulness of IV-technique is presented by Newhouse and McClellan (1998) who explain the instrumental variable convincingly as a device that achieves a pseudo-randomization. The authors analyzed the effect of catheterization and associated revascularization of acute myocardial infarction on mortality in the years following treatment. For IV-estimation of this effect, they used the differential distances of the patients' place of residence to the (nearest) catheterization, revascularization, and high-volume hospital as instrumental variables.

24.3 Demand, Need, Utilization, and Access to Health Care

A main focus of HSR is the assessment of demand, need, utilization, and access to health services, which represent closely related but distinct fields of investigation. In the input-output model, they represent the endogenous, risk-related input, which is, among others, determined by population health – representing the exogenous risk-related input in the model.

Demand is a general economic concept, which can be defined as the "quantity of a good buyers wish to buy at a conceivable price" (Begg et al. 1997a). Demand for health and health care is in many respects different from demand for other goods and services. The demand for health care is a derived demand as health care is not sought in itself but as a means to improve one's health or to prevent its deterioration. Health care itself is indeed often rather unpleasant (McPake et al. 2002). Health is not something that can be traded, and both health and health care are surrounded by

uncertainty. What people want in essence from health care is to buy access to care in case they need it, that is, insurance (McPake et al. 2002). Another aspect is that health is both a consumption good and a capital good (McPake et al. 2002). Especially politicians and health-care funders often focus on the consumption side and neglect the potential of investing in health as a durable good which is an important prerequisite for economic growth.

The notions of need, utilization, access, and the relationships between them will be discussed in more detail in the following sections.

24.3.1 Assessing Health Needs

The concept of "need" for health and health care links directly to population health. Initially it appears simple and is often used by politicians in health policy discussions but quickly becomes complicated and is therefore avoided by many analysts (Kindig 1997). Instead many policy analysts, in particular, in the United States, prefer to use the economic demand and supply framework, where it is assumed that if someone "needs" something, he or she will express this desire by purchasing the item that is needed in the marketplaces, and as a consequence, supply will increase (Kindig 1997).

An alternative concept of need as the "capacity to benefit" has been proposed by Williams (1974) and Culyer (1993). Their concept also goes beyond the perspective of an initial baseline level of health because unhealthy individuals and populations cannot be said to need more health care without regard to their potential for improving their health status (Kindig 1997). Capacity to benefit also rules out health services which might be desired by individuals or providers but which do not make a positive contribution to health-adjusted life expectancy (Kindig 1997). It also goes beyond a mere epidemiological description of health needs in terms of ill-health or shortcomings in care in a specified population as it incorporates the notion of effectiveness of the intervention. Some authors use the terms "felt need," "normative need," and "expressed need." The "felt need" reported by patients is often substantially different from the "normative need" as judged by health professionals. "Expressed need" represents the need expressed by action, for example, visiting a doctor (Wright 2001).

Three main approaches to health needs assessment exist (Wright 2001):

- Epidemiologically based needs assessment – combining epidemiological approaches, such as specific health status assessments, with assessment of the effectiveness and possibly cost-effectiveness of interventions
- Comparative needs assessment – comparing levels of service receipt between different populations
- Corporate needs assessment – canvassing the demands and wishes of professionals, patients, politicians, and other stakeholders

In practice, comprehensive health needs assessments often combine all three approaches. Practical applications are manifold, such as to highlight areas of unmet

need and to provide a clear set of objectives to meet these needs, to decide how to use resources to improve the local population's health, and to influence policy, interagency collaboration, or research and development priority setting (Wright 2001).

A good example for a health needs assessment is a study contrasting the epidemiological need for carotid endarterectomy with actual service provision in an English region (Ferris et al. 1998). The authors estimated the need for a carotid endarterectomy on the basis of demographic and epidemiological data, assuming that the rate of endarterectomies in their region should match the rate of patients with symptomatic carotid disease – the patient group for whom carotid endarterectomy is proven to prevent strokes. Based on estimates of the incidence of transient ischemic attacks ($77/10^6$ population/year) and minor strokes ($76/10^6$/year), they calculated that the need for endarterectomy was $153/10^6$/year, which contrasted with operation rates of $35/10^6$/year in 1991–1992 and $89/10^6$/year in 1995–1996. The ratio of use to need was 0.47 (95% confidence interval (CI) from 0.4 to 0.54), which was far from being satisfactory. Furthermore, they noted a disconcerting variation in the use to need ratios between districts, ranging from 0.28 (95% CI: 0.19–0.38) to 0.81 (95% CI: 0.62–1.06), and lower use to need ratios in elderly and female patients – indicative of inequity in access in relation to need. The epidemiological needs assessment was supplemented with a corporate needs assessment comprising interviews with vascular surgeons and a joint purchaser-provider workshop. These indicated that the low operation rates were primarily due to low rates of referral for diagnostic assessment by general practitioners. The variation between districts partly reflected the concentration of services – districts with a high use to need ratio tended to have one of the main provider sites. This study clearly demonstrates the usefulness of such research in identifying the main levers for improvement of current service provision – in this case, raising awareness for the clinical indications for carotid endarterectomy in general practitioners, in particular, those located in rural areas without access to a local vascular surgical service.

24.3.2 Assessing Utilization and Access to Services

Studies assessing the utilization of services attempt to improve our understanding of who uses health-care services and why (Black 1997). In the previous sections, it has already become clear that many factors determine whether a patient utilizes a health-care service, among them whether the patient suffers from a condition for which an effective intervention is available and whether he or she demands that service. Three other common determinants of utilization have become apparent in the example of the health needs assessment for carotid endarterectomy in England – clinician's judgment, distance from facilities, and sex. If a general practitioner does not consider that referral to a specialist is necessary, it is unlikely that the patient will end up having the procedure nonetheless – resulting in unmet need and underutilization of health services. Other important factors influencing utilization and access are patients' knowledge and the cost of services to the patient.

A well-recognized example for the influence of sex on utilization is higher rates of appendectomy in women than in men (Black 1997). After the primary assertion that appendicitis is more frequent in women was not supported by evidence, the more likely explanations were as follows: Appendicitis-like symptoms are more common in women, probably arising from ovarian dysfunction; in some cultures, young women prove their independence by undergoing an operation; operation rates are dependent on the availability of services – the more surgical services are available, the higher the sex difference in operation rates (Black 1997).

The influence of clinicians' judgment on health service delivery has been investigated in a series of studies comparing hospital care and related costs between Boston and New Haven in the United States – two cities with similar demographics where most hospital care is provided by university hospitals (Wennberg et al. 1987, 1989). However, in 1982, expenditures per head for inpatient care in Boston were about twice as high as in New Haven ($889 vs. $451) (Wennberg et al. 1987). The excess utilization in Boston compared to New Haven totaled $300 million and 739 hospital beds per year. In a subsequent study in 1985, Wennberg and colleagues showed that the variation in operation rates between the two areas did not result in statistically significant differences in mortality between the two cities (Wennberg et al. 1989). The authors concluded that the lower rate of hospital use in New Haven was not associated with a higher overall mortality rate in the populations concerned and consequently that hospital care was overutilized in Boston and not underutilized in New Haven.

The influence of geographical and financial barriers to access is also well documented. Black and colleagues (Black et al. 1995) attempted to identify the reasons for geographical variation in the use of coronary revascularization in the United Kingdom in a cross-sectional study. They found considerable variation in revascularization rates between districts, which arose from differences in supply factors, notably the distance to a regional revascularization center and the existence of a local cardiologist. The level of coronary heart disease mortality in the population and the lack of use of alternative treatments not only failed to explain the observed variation but was inversely associated with the rate of revascularization (Black et al. 1995). This inverse relationship between need and provision of care has been observed in many settings and has been termed the "inverse care law" by Julian Tudor-Hart (Tudor-Hart 1971, 2000). It has to be kept in mind that in measuring the utilization of health-care facilities, only those patients are counted, who have surmounted barriers to access – be it long distances, fear of an operation, lack of public transport, waiting lists – and are thus biased (Schwartz and Busse 2003). These barriers to access also exist in countries which grant a legal right to health care to every citizen, in particular, among socially disadvantaged groups in society. These differences in access to care are even more pronounced in countries without such a right to health care and where direct financial barriers to care exist on top of other barriers to access.

The effect of financial barriers to accessing health services has been studied in many countries at different levels of development. The most famous study is the RAND Health Insurance Study (see also Sect. 24.2.1.1) on the effect of cost-sharing

measures on utilization (Newhouse 1974; Newhouse et al. 1981; Brook et al. 1983). Between 1974 and 1977, about 2,000 non-elderly families were randomly assigned to different insurance plans. Participants were assigned to either prepaid group practices or to one of 14 fee-for-service insurance plans, which varied in their coinsurance rates and in maximum spending per family and year (Newhouse 1993). The authors found that the more families had to pay out of pocket, the fewer health-care services they used. Families on the plan with the highest coinsurance (95%) up to a $1,000 limit on annual family expenditure reduced expenditure by 25–30% compared to a plan which was free to the family (Newhouse 1993). Interestingly, the use of all types of services, whether physicians, hospitals, pharmaceuticals, dental, or mental health services, fell with cost-sharing to a similar degree, except hospital admissions of children which did not respond to plan (Newhouse 1993). While the reduced utilization had no negative effect on the health for the average person, health among the "sick poor" – the most disadvantaged 6% of the population – was adversely affected (Newhouse 1993). Especially the poor who began the experiment with elevated blood pressure had their blood pressures lowered more on the free care plan than on the cost-sharing plans, and mortality rates predicted on the presence of major risk factors were 10% lower among those insured on the free plan (Newhouse 1993). Free care at the point of delivery also improved both near and far corrected vision, increased the likelihood that a decayed tooth would be filled, and the prevalence of anemia among poor children was lower (Newhouse 1993). All observed adverse health effects of cost-sharing hit the poor and less educated disproportionately more.

A number of other factors can limit access to services, in particular, gender, age, professional status, race, and religion (Schwartz and Busse 2003). The discrepancy between need in terms of ill-health and capacity to benefit from intervention and utilization is a commonly used measure of equity.

24.4 Financial Resources, Structure, and Organization of Health Services

Health-care financing can be described from different perspectives:

- The first one looks for the ultimate sources of funding. Here, intermediary sources of financing (government, social security funds, private social insurance, and private households above all) have to be tracked back to their origin.
- The second one, commonly used in National Health Accounts, aims at a breakdown of expenditure on health into the complex network of third-party payments plus the direct payments by households or direct funders (OECD 2000).
- The third one focuses on the allocation of the available resources. Health planning and management of health care among others includes the continuous task of distributing the financial resources, for example, to distinct segments in the natural history of disease (as reflected in prevention, cure, rehabilitation, and care), to alternative treatments for a specific disease, to different regions, to various groups of providers, or simultaneously with respect to some of these categories.

When comparing the developed countries with respect to financing from the first perspective, we can find two marked types of health services systems: systems that are funded by taxes and typically have a National Health Service (NHS) – so-called NHS-type or Beveridge-type systems – and systems that are predominantly financed by contributions of employees and employers (so-called Bismarck-type systems). In the real world, we find, of course, mixed systems including all the common forms of public and private financing. The UK is generally considered to be the classic example of an NHS-type health service system, and the German health system, as established by Bismarck in the late nineteenth century, naturally acts as the model of all Bismarck-type systems.

Health service systems also vary considerably in the overall health insurance coverage rates. In all EU member states, nearly 100% of the population has coverage of some type of public or private health insurance, whereas in the USA, 46.3 million persons (corresponding to 15.4% of the population) had no coverage at all during the entire year 2008 (U.S. Census Bureau 2009), which constitutes an important barrier to access as seen above.

International comparisons of health-care spending are hampered by a variety of definitions and classifications, used by the national statistical agencies, resembling the different organizational structures of health-care delivery (van Mosseveld 2003). To improve comparability of health, accounting data Eurostat, the statistical office of the European Union (EU), and the Organization for Economic Co-operation and Development (OECD) jointly developed a conceptual basis of rules for the statistical reporting of health accounts together with EU member states (OECD 2000). This so-called *System of Health Accounts (SHA)* corresponds to the new German health expenditure data system which was developed simultaneously (Brückner 1997; Statistisches Bundesamt 2000a, b). The SHA includes expenditures of private households, is consistent with the System of National Accounts (SNA) methodology (OECD 1993), and covers three dimensions: functions of health care, providers of health care, and sources of funding. Core of the SHA is a newly developed *International Classification for Health Accounts (ICHA)* that is based on a three-digit code.

The ICHA reflects the potential variations among health systems in structure and organization of health care and the share of work between the various providers. One remarkable distinction between health services systems, when looking at the organization, refers to outpatient care. In many countries, patient-physician contacts take place only in offices of general practice. Contacts with medical specialists are limited to hospital visits no matter whether the patient has been admitted to the hospital or not. But there are other countries, among them Germany, where outpatient specialized curative care is predominantly provided by office-based specialists.

There are many additional differences in the structure and organization of health services systems in spite of the fact that in all developed countries, patients with a specific health condition receive more or less the same treatment, provided that quality standards are observed. The package of activities in health care seems to be stable over the health systems, while the providers are different. International

studies of health-care systems based on comparable data sets are rare. One of the outstanding examples is the WHO/International Collaborative Study of Medical Care Utilization – WHO/ICS-MCU (Kohn and White 1976). The data collection process of this carefully designed and methodically ambitious study included a cross-national multilingual household survey (by personal interviews) and standardized forms relating to health services resources and organizational factors. The study included almost 48,000 respondents representing over 15 million persons in 12 study areas scattered over Europe and America. One of the striking results of the WHO/ICS-MCU study was that study areas with the highest estimates of societal interest in health were also the areas with the lowest totals for per capita health expenditure and for health expenditure as a percentage of national income.

24.4.1 Allocation of Resources

Health-related decision makers in government, regional authorities, insurance companies, or other institutions are faced with the task of allocating resources. Examples are allocating research funds to different areas of HSR, Medicaid funds to treatments, Medicare funds to HMOs, or global funds to local authorities. Regional allocation is a main concern of all NHS-type health-care systems. Risk-adjusted capitation payments to insurers or to providers is a related topic that has been discussed extensively in several non-NHS-type countries (e.g., the USA, the Netherlands, and Germany).

There is no unique method for resource allocation analysis. Different levels and variable purposes of resource allocation analysis require different methods:

- For economic evaluation, for example, of drugs, surgical procedures, other types of clinical interventions, or of community intervention programs, limits on health-care resources mandates resource-allocation decisions guided by considerations of cost in relation to expected benefits (Weinstein and Stason 1977).
- The UK approach of weighted capitation has become the principal method of allocating health-care finance to regions (Rice and Smith 1999).
- Risk-adjusted capitation, whereby capitated payments are adjusted to reflect the expected cost of individual enrollees, is commonly based on multivariate regression models to predict health-care expenditure (Van de Ven and Ellis 2000).

While economic evaluation are mostly based on RCTs and observational studies (mainly on effectiveness and costs), risk-adjusted capitated payments and formulas of weighted capitation are generally based on official statistics (e.g., census or mortality statistics) or large samples from administrative databases. For an example, see Sect. 24.4.1.2.

24.4.1.1 Economic Evaluation, Especially Cost-Effectiveness Analysis

Economic evaluation can be defined as the comparative analysis of alternative courses of action in terms of both their costs and consequences (Drummond et al. 1997). There are four main types of economic evaluation (see Table 24.3).

Table 24.3 Types of health economic evaluations (Source: Epstein and Sherwood (1996))

Type of analysis	Assumption/question addressed
Cost-minimization	The effectiveness (or outcome) of two or more interventions is the same. Which intervention is the least costly?
Cost-effectiveness	The effectiveness of two or more interventions differs. What is the comparative cost per unit of outcome for the intervention?
Cost-utility	The question is the same as for cost-effectiveness analysis. The outcome is a preference measure that reflects the value patients or society places on the outcome.
Cost-benefit	The effectiveness (or outcome) of two or more interventions differs. What is the economic trade-off between interventions when all of the costs and benefits of the intervention and its outcome are measured in monetary terms?

But in practice, most of the health economic evaluations apply the cost-effectiveness methodology. *Cost-benefit analysis* is rarely used in public health and health-care settings because of methodological difficulties to measure the value of human life and low acceptability of its results on the side of health policy decision-makers and health professionals. *Cost-utility analysis (CUA)* is considered by many to be a subtype of *cost-effectiveness analysis (CEA)* where the effectiveness measure includes societal or individual preferences for the outcomes – a customary effectiveness measure in CUA is the quality-adjusted life year (QALY) (cf. Sect. 24.6.1.2 and chapter ▸Descriptive Studies of this handbook) – compared to natural units as effectiveness measure in CEA, for example, life years gained or mmHg blood pressure reduction. Most of the important methods and concepts applicable to cost-effectiveness studies are also applicable to cost-utility and cost-minimization studies. Cost-of-illness studies may be identified as a fifth type of economic study in HSR. Their goal is to estimate the total societal costs of caring for persons with a specific illness compared to persons without this illness, irrespective of any intervention. Such studies are carried out to demonstrate the (relative) burden of illness. They are not full economic evaluations because alternatives are not compared (Drummond et al. 1997).

One limitation that is common to all types of economic evaluation arises from the difficulty in obtaining a true estimate of costs, particularly in a health-care or public health setting where high proportions of fixed costs and little flexibility in changing the labor pool are typically found (Petitti 1998b).

A common understanding of cost-effectiveness claims that one of the three criteria has to be met (Doubilet et al. 1986). First, an intervention is cost-effective when it is less costly and at least as effective as its alternative. Second, an intervention is cost-effective when it is more effective and more costly, but the added benefit is "worth" the added cost. Third, an intervention is cost-effective when it is less effective and less costly, and the added benefit of the alternative is not "worth" the added cost.

Cost-effectiveness is measured as a ratio of cost to effectiveness. Two concepts to calculate this ratio should be distinguished (Detsky and Naglie 1990): An average

cost-effectiveness ratio is estimated by dividing the cost of the intervention by a measure of effectiveness. An incremental cost-effectiveness ratio is an estimate of the cost per unit of effectiveness of switching from one intervention to another. In estimating an incremental cost-effectiveness ratio, both the numerator and denominator of the ratio represent differences between alternative interventions (Weinstein and Stason 1977). Often the terms "marginal" and "incremental" are used interchangeably in the literature, although *marginal costs* are strictly speaking the costs of producing one extra unit of output, whereas *incremental costs* usually refer to the difference, in cost or effect, between the two or more programs being compared in the economic evaluation (Drummond et al. 1997).

Estimating *average cost-effectiveness ratios* can be useful for service planning and for resource allocation decisions between very different health programs, for example, an influenza vaccination program and liver transplantations. However, for resource allocation decisions between interventions for the same disease, for example, two different antihypertensive drugs, *incremental cost-effectiveness ratios* should be used. The importance of using incremental cost-effectiveness ratios for decision making in some settings is best illustrated with the example of the sixth Guaiac stool test to screen for colorectal cancer, which had been endorsed by the American Cancer Society and which has later been shown to have an incremental cost of $47 million per case detected compared to an average cost of $2,451 per case detected (Neuhauser and Lewicki 1975).

The unspecified implicit alternative to an intervention is usually doing nothing. But doing nothing has costs and effects that should be taken into account in the analysis (Detsky and Naglie 1990). Furthermore, explicit declaration of "doing nothing" as the alternative intervention helps to frame discussions of the desirability of the intervention (Petitti 1998b).

Costs seem to be a straightforward notion, well understood by everybody. But actually, it is a rather complex term that consists of various components: direct costs, indirect costs, and intangible costs. Costs that are directly related to an intervention (and to side effects and other consequences) are summed up to the total of *direct costs*. By *indirect costs,* health economists understand the monetary value of lost wages and productivity due to morbidity and death of a person affected. *Intangible costs* refer to consequences that are difficult to measure and value, such as the value of improved health or the pain and suffering associated with illness or treatment (Drummond et al. 1997). The rationale of economic evaluation is based on the concept of *opportunity cost*, that is, the benefits forgone by not deploying resources for the next best alternative use.

As costs are seen differently from different *perspectives* (e.g., perspectives of health insurers, corporations, hospitals, physicians, and patients), it is also important to define a cost perspective in CEA and state it explicitly (Petitti 1998b). A common goal in CEA is the societal perspective so that the total costs of the intervention to all payers for all persons are included in the analysis.

Costs and benefits, after all, must be discounted before comparing them by calculating the ratio of cost to effectiveness. *Discounting* is the usual procedure in economics used to determine the present value of future money. This analysis gives

a greater weight to costs and benefits the earlier they occur. High positive discount rates favor alternatives with costs that occur later or benefits that occur earlier. This clearly favors curative versus preventive health programs. In the business world, there is no fixed rate of return on investment, and the use of a private sector return rate for public sector program cost may not be correct (Sudgen and Williams 1990). Most published CEA in developed countries use discount rates between 3% and 5%. An expert panel commissioned by the US Public Health Service, based on the "shadow price of capital," recommended using a discount rate of 3% for economic evaluation in the public health sector (Gold et al. 1996). Whether benefits should also be discounted, and if so at what rate, is highly controversial.

Estimates of benefits and costs in a CEA may be uncertain because of imprecision in both underlying data and modeling assumptions. Therefore major assumptions should be varied and the net present value and other outcomes computed repeatedly to determine how sensitive outcomes are to changes in the assumptions. This so-called *sensitivity analysis* is typically the last step in a CEA. A sensitivity analysis varying the discount rate from 0% up to 7% should always be done (Gold et al. 1996).

As illustrated by Oregon's Medicaid reform efforts in 1990/1991, CEA or other types of economic evaluation cannot be used as sole basis for allocating scarce resources because the question of equity and ethical issues are not addressed by this method. In Oregon, CEA was only 1 of the 13 factors used to prioritize funding of services for the poor (Petitti 1998b).

Since Sir William Petty found out in 1667 that public health expenditures to combat the plague would achieve a benefit-cost ratio of 84 to 1 (Fein 1971), numerous studies of economic evaluation have been carried out, most of them using ratios of cost to effectiveness. The list of interventions that were economically evaluated within the last 10 year spans from influenza vaccination of healthy school-aged children (White et al. 1999) to colonoscopy in screening for colorectal cancer (Sonnenberg and Delco 2002), and preoperative autologous blood donation (Etchason et al. 1995) to reducing the population's intake of salt (Selmer et al. 2000).

24.4.1.2 Weighted Capitation in NHS-Type Health Systems

The central aim of weighted capitation is to distribute a global health budget between geographical areas in accordance with population needs and thus provide equal opportunity of access for equal needs. Currently used formulas of weighted capitation can be described as a modified age standardization of health-care expenditure. The UK Resource Allocation Working Party (RAWP) originally recommended in 1976 that the resources for the hospital and community health services (HCHS) be distributed on the basis of population size, weighted by age and sex, the need for health care, and the costs of providing services (Carr-Hill 1989; Advisory Committee on Resource Allocation (ACRA 1999)). Standardized mortality ratios (SMRs) were used as a proxy measure for relative needs. However, this had been criticized for failing to fully reflect the demand for health-care resources related to chronic disease and deprivation.

In 1995, a new weighted capitation formula for HCHS was introduced. This comprises an age index (based on estimates of national resources spent per capita in eight age groups) and an "additional need" index (additional to that accounted for by demographic variables). The need weighting index takes the form of four indices for acute, psychiatric, non-psychiatric community, and community psychiatric services, which are based on 1991 small-area census socioeconomic variables. It is derived from an empirical model that identified its need indicators as those census-derived health status and socioeconomic variables which, having been adjusted for the independent effects of supply, were most closely correlated with the national average pattern of hospitalization (Carr-Hill et al. 1994).

For all its merits, however, this formula, also called English formula, and the models on which it is empirically based have been criticized. The fundamental criticism relates to the use of utilization-based models to assess need for health care, which implies that historical patterns of service uptake between different care groups (as revealed by utilization) are appropriate (Mays 1995).

Against this background, some scientists pleaded for a radically new approach to health resource allocation, one that distributes NHS resources on the basis of direct measures of morbidity rather than indirect proxies such as health service utilization or deprivation. The Welsh steering group on allocation (Townsend 2001), for example, recommended the use of a morbidity-based budgeting approach. In a study of target allocations for the inpatient treatment of coronary heart disease in a sample of 34 primary care trusts in different areas in England, it was shown that a morbidity-based model would result in a significant shift in hospital resources away from deprived areas, towards areas with older demographic profiles and toward rural areas (Asthana et al. 2004). In the discussion of their findings, the authors concluded by calling for greater clarity between the goals of health-care equity and health equity.

Up from the year 1999, the Advisory Committee on Resource Allocation (ACRA) of the Department of Health took care of the further development of the Weighted Capitation Formula. ACRA picked up the above mentioned critique and incitations and learnt to distinguish between the two objectives "equal access to healthcare" and "reduction of health inequalities." Moreover, an index to estimate unavoidable labor costs, known as the staff Market Forces Factor (MFF), was included in the formula.

The actual formula for 2009–2010 and 2010–2011 allocations is built up by three components: hospital and community health service, prescribing and primary medical services. The corresponding weights are 76%, 12%, and 11%. Each component reflects both the additional needs and the unavoidable costs (the prescribing component does not have an adjustment for unavoidable costs since the prices of drugs do not vary by geographical location.)

The components again are set up by two indices: one (model-based computed) aiming at utilization of services and the other one aiming at health inequalities uses disability-free life expectancy as its measure, combining 2005 life expectancy data with 2001 limiting long-term illness data, and so capturing morbidity as well as mortality (Department of Health 2008).

24.4.1.3 Risk-Adjusted Capitation

When competition is an essential component of a health-care system, it is a widespread belief that capitated payments create incentives to contain costs and to compete on quality. But they also create undesirable incentives for risk selection ("cream skimming"), that is, to attract profitable patients (or enrollees) and to avoid unprofitable ones, and to decrease service intensity.

Risk adjustment is an important tool to reduce cream skimming while encouraging desirable cost and quality competition. This method controls for confounding (comorbidity above all) by calculating the expected health-care costs (or some other measure of an outcome) for members of health plans or insurance companies. This control is realized either by stratification (*cell-based approach*) or by multivariate modeling (*regression approach*). In many developed countries around the world, health-care organizations have established some sort of risk adjustment procedure for resource allocation. Examples exist for the following countries: Austria, Brazil, Canada, Chile, Germany, Hong Kong, Israel, the Netherlands, Spain, Switzerland, Taiwan, and the USA. Most of the risk adjustment procedures are based on a regression model to predict future health expenditure. Models differ with respect to the set of included predictors, the procedure of grouping diagnostic information, and the populations used for calibration.

The exclusive reliance on risk-adjusted capitated payments has been criticized, for example by Newhouse et al. (1997), who pointed out that the common risk adjusters (predictors of cost) are not likely to reduce risk selection problems to negligible levels. This concern was confirmed by a study of Shen and Ellis (2002) who examined the maximum potential profit that plans could hypothetically gain by using their own private information to select low-cost enrollees when payments are made using one of four commonly used risk adjustment models. Their findings – based on simulations using a privately injured sample – suggested that risk selection profits remain substantial (Shen and Ellis 2002).

Against this background, it was recommended to move the financing of health services to partial capitation payments. Partial capitation for an individual enrollee combines capitation methods and some reflection of that person's actual use of services, for example, a fee for service payment. Partial capitation would reduce plans' incentives to select good risks – the intent of risk adjustment – and also reduce the financial incentive to underserve or stint on care (Newhouse et al. 1997).

The General Form of Regression-Based Risk Adjustment Model Frequently used are regression models with untransformed costs as the dependent variable, estimated by ordinary least square (OLS). The standard assumptions of that type of statistical model (namely, a normal distribution, homoscedasticity, and independent observations) are not satisfied sufficiently by utilization data, but for predicting future costs, the model has shown to work about as well as more complex models in real situations (Diehr et al. 1999).

Occasionally a two-part model is applied: One equation predicts the probability that a person has any use, and a second equation predicts (on a log scale) the level of use for users only. In a two-part model, the regression coefficients of the first equation are estimated by logistic regression analysis and those of the second equation by OLS regression. Two-part models tend to meet the assumptions better than one-part models and provide insight into the utilization process, but they are not recommended when the goal is to predict future costs because transformations cause complications in this context (Diehr et al. 1999).

The list of possible predictors of a model for risk-adjusted capitation includes age, sex, and other demographic or socioeconomic variables, as well as binary variables to indicate that a person has been assigned to a diagnosis belonging to a special group from a system of diagnostic groups or has received a drug prescription belonging to a special group from a system of drug categories. To incorporate information on morbidity, some models use hospital diagnoses alone, while others use both inpatient and outpatient diagnoses. It has, however, to be noted that previous utilization is a strong predictor of a future one. This means that the costs in 1 year heavily depend on utilization in the year before for chronically ill patients. Thus, as in any regression analysis, it is important not to control for this variable lying on the causal pathway (Diehr et al. 1999).

The estimated regression coefficients ("regression weights") refer to the so-called calibration population. For diagnosis-based models, generally this is also the population used to establish the diagnostic classification system, the "grouper." Recalibration of a model without a refinement of the grouper therefore may lead to biased estimation. Generally the models are calibrated prospectively (that is, the data of the predictor set refers to the previous year, while the cost data refer to the actual year), but in order to evaluate the predictive power, current calibration (both types of data refer to the same year) has been performed as well.

The standard summary measure of model performance in prediction is R^2, the percentage of the total variance of the dependent variable that is explained by the model. Usually the values of R^2 in prospectively calibrated models do not exceed 20%. Newhouse et al. (1989) used theoretical and empirical arguments to estimate that the maximum possible R^2 in the context of utilization data is about 15% for total expenditure (prospectively modeling).

In addition to the grouper and the regression module, any risk adjustment methodology finally requires a module that links the estimated costs to the payment system or the resource allocation procedure, respectively, that is, a mechanism that controls the way payments or the allocation of resources is based on the predicted health-care expenditure.

Risk Adjustment in the US Setting Up to 1999, Medicare paid the HMOs 95% of the *adjusted average per capita cost (AAPCC)*, an estimate of the expected cost of treating Medicare beneficiaries in the fee-for-service sector in each local area. The AAPCC methodology adjusted for differences between the HMOs enrollees and fee-for-service users with respect to age, sex, welfare status, and whether or not they were in a nursing home (Ellis et al. 1996).

Since its implementation in 1985, the AAPCC had prompted concern about its fairness and accuracy, and it was shown that only about 1% of total variance of the cost of treatment was explained by this concept (Newhouse 1986; Ash et al. 1989). Against this background, the *Health Care Financing Agency (HCFA)* sponsored the development of alternative approaches that include diagnostic information as predictors in the regression-based risk adjustment model, among them the *Diagnostic Cost Groups (DCG)* family and the *Adjusted Clinical Group (ACG)* methodology (Ingber 1998). In the years 2000–2003, AAPCC has been stepwise replaced by the *Principal Inpatient Diagnostic Cost Group* model *(PIP-DCG)* which uses sociodemographic variables and hospital diagnoses to predict next years cost (Pope et al. 2000). In view of the widespread concern about the quality of ambulatory diagnoses, the DCG family was supplemented in 2001 by a model that uses outpatient pharmacy data, grouped into 127 mutually exclusive categories, instead of ambulatory diagnoses (Zhao et al. 2001). From 2004 onward, the CMS/HCC-model,[7] a 100% comprehensive risk adjustment scheme (using full encounter diagnostic data) has been implemented to adjust Medicare capitation payments to private health-care plans for the health expenditure risk of their enrollees (Pope et al. 2004).

Medicaid supported the development of the *Chronic Illness and Disability Payment System (CPDS)* which groups the Medicaid beneficiaries according to a hierarchical diagnostic classification system (Kronick et al. 1996). CPDS, which later on was reconstructed and recalibrated to predict expenditures also for Medicare beneficiaries, has now been established in several US states (Kronick et al. 2002).

Risk Adjustment in a Bismarck-Type European Setting In European countries with a predominating Bismarck-type organization of health services, we find competition among all insurance companies and even among the statutory sickness funds. The main goal of risk adjustment (better: risk equalization) in these settings therefore is to reduce risk selection by the sickness funds and to establish a fair system of income-related contributions. HSR has played a major role in designing and reforming these systems.

Like in the USA, the starting point of risk adjustment in Europe has been set by models based on age, sex, and other sociodemographic variables. The Netherlands, for example, started in 1992 with a prospectively used age- and sex-based model. In 1995, region and disability were included as predictors, and a "high-risk pool" was established in addition. Since 2002, dummy variables were added to the model that indicate prescriptions of drugs falling into 1 out of 13 mutually exclusive categories, the *Pharmacy-based Cost Groups (PCGs)*, each of them closely related to a serious chronic disease (Lamers 1999). From 2004 onward, the Dutch risk-equalization methodology had been further supplemented by an inpatient DCG module that uses hospital diagnoses only.

[7]CMS: Centers for Medicare & Medicaid Services; HCC: Hierarchical Condition Categories

In 1994, Germany introduced a retrospective risk-equalization procedure among statutory sickness funds which was based on the following variables: age, sex, and two dummy variables indicating invalidity or disability pension and the entitlement for sickness allowance. The procedure was also designed to adjust for different incomes because the beneficiaries pay income-related contributions. The largest share of the risk-adjusted financial transfers between sickness funds (up to 60%) results from differences in per capita income of the beneficiaries. From 2002 onward, the German risk-equalization methodology has been extended. First, a retrospective "high-cost pool"[8] was established, and second, a dummy variable was added to the set of risk adjusters indicating that a beneficiary is registered in an accredited disease management program. From 2009 onward, a risk-equalization procedure based on a DCG/HCC module has been established, using inpatient and outpatient diagnoses with respect to 80 (by an expert panel) selected diseases.

24.4.2 Evaluating Effects of Organizational Characteristics and Change

As health services systems in the developed countries tend to go through one reform after another and are more or less continuously exposed to change, evaluation is a permanent task of HSR. But the preconditions do not favor the establishment of scientifically sound designs of research. Experimental designs are extremely rare. The above-cited RAND Health Insurance Study on the effect of cost-sharing measures on utilization is one of the most famous exceptions. In some circumstances, it is even difficult to implement a quasi-experimental design including a control group. Particularly in countries like Germany, where benefits and programs are uniform but the organizational responsibilities are widely scattered over local authorities and institutions, evaluation research is very complex.

Suggested by the structure of available data, perhaps the most frequently used quasi-experimental design for analyzing aggregated annual data in the context of program evaluation is the time series experiment. It can be characterized by a periodic measurement process on some group and the introduction of an experimental change X into this time series of measurements O_i. Adapting a diagram by Campell and Stanley (1966), the time series design can be outlined as follows (whereby the number of observations before or after X, occurring here in year five, may be smaller or larger as in a real problems):

$$O_1, O_2, O_3, O_4, X, O_6, O_7, O_8, O_9$$

The main problem of (internal) validity inherent in a time series design is revealed by seeking likely alternative explanations of the shift in the time series

[8]The high-cost pool consists of insured with high cost in the past year (above a fixed threshold) which are shared by all statutory sickness funds.

other than the effect of X. This problem, of course, could be settled to a great extent by establishing a suitable control group (comparison series) that shares all intervening factors except X with the study group.

A natural approach for analyzing data from a time series design is *segmented* or *piecemeal regression* (e.g., Neter and Wasserman 1974). This method is appropriate when the considered response variable has a linear trend over the range before X (segment one) followed by another linear trend over the range after X (segment two). The year which divides the segments (year five in the above diagram) is known as the join point (or break point). When the hypothetical change of trend line refers only to the slope and not to the intercepts (that means no discontinuity between the both lines), the regression equation for analyzing data from a design as diagrammed above can be specified as follows:

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_1 x_2,$$

where Y is the response variable, x_1 is the year ($x_1 = 1, 2, 3, 4, 6, 7, 8, 9$) and x_2 is a dummy variable indicating that the year is greater than 5. The parameter β_2 measures the difference in slopes between the lines. If the same trend continues from the first segment to the second segment, then $\beta_2 = 0$. The β_j are estimated, and the hypothesis $\beta_2 = 0$ is tested, using standard procedures in regression. It is easy to expand segmented regression to more than two segments and even to allow for discontinuities between segmented regression lines. Autocorrelated errors or heteroskedasticity can be handled by using standard techniques (e.g., Greene 2003).

A good example for applying this type of analysis to evaluate the impact of a program on the basis of aggregated data is the study of the effect of a regionalized perinatal care program in North Carolina (established in 1965) on perinatal and postneonatal mortality (Gillings et al. 1981). A similar – but due to autocorrelation, more complex – analysis was carried out to evaluate the effects of patient-level payment restrictions for prescription drugs under Medicaid in the years 1981–1983. By this analysis, supplemented by survival analysis to measure the rate of admissions to hospital and nursing homes, it could be shown that the decline in the use of drugs after the cap (a limit of three paid prescriptions per month) had been associated with an increase in rates of admission to nursing homes (Soumerai et al. 1987, 1991).

Regardless of these examples, there is only limited use of OLS regression models for evaluation because of several restrictions. First, costs or the logarithm of costs or other continuously distributed responses are only one type of outcome measure used for evaluation. Counts of specific events, for example, contacts, prescriptions, hospital admissions, etc., and binary response variables like death, accident, or first occurrence of a specific disease are equally important – in some situations, even more important measures. Second, if individual longitudinal data are available, to make full use of the data structure, the model used should be able to handle clustered (correlated) response data, arising from repeated measurements and time-varying covariates. *Generalized linear models* (Nelder and Wedderburn 1972;

McCullagh and Nelder 1983) and the related *generalized estimating equations* (GEEs) (Liang and Zeger 1986; see also chapter ▸Generalized Estimating Equations of this handbook) form the methodological framework of an advanced approach of statistical modeling for evaluation. Consistent parameter estimates in these models are achieved by maximizing likelihood or quasi-likelihood functions using some sort of Gauss-Newton algorithm. Several of the common packages of statistical software, among them SAS, provide corresponding procedures.

The standard model to analyze count data, for example, is the *Poisson regression model*, which is a non-linear regression model that can be formulated as a generalized linear model. Poisson regression is robust insofar as consistent estimation of the regression coefficients does not require that the dependent variable is Poisson distributed. Only a correct specification of the conditional mean is required (Cameron and Trivedi 1998). But Poisson regression is prone to overdispersion. Therefore the condition that the variance equals the mean has to be relaxed by introducing a dispersion parameter that must be estimated as well. Otherwise, testing hypotheses on the regression coefficients could yield misleading rejections of null hypotheses (Cameron and Trivedi 1998).

Poisson regression can also be used to analyze correlated counts from repeated measurements. The within patient correlation is then estimated in the framework of GEEs, whereas the effects of the covariates can be modeled as a generalized linear model. For example, the introduction of reference pricing for angiotensin-converting enzyme (ACE) inhibitors for patients of 65 years of age or older in British Columbia, Canada, in January 1997, was evaluated by using such a type of analysis (Schneeweiss et al. 2002). Several covariates were included in the model, among them age, sex, the adjusted household income, and a chronic disease score computed from prescription medications for every quarter and treated as a time-varying covariate. This ambitious study was based on computerized administrative health databases covering a large proportion of the population, including all types of claims, hospital admissions, admissions for long-term care, diagnoses, and the medications, dose, and dispensed quantity of all prescriptions. Even the deaths within the study cohort were included.

A similar analysis has never been done in Germany, though reference-based prices (RBP) for the beneficiaries of the statutory sickness funds were established in 1992/1993. For reasons of privacy and data protection, cross-institutional linkage of existing scattered administrative databases on drug utilization, ambulatory diagnoses and medical services, and hospital data on an individual level need extensive data protection procedures in Germany. Thus the effects of RBP has to be evaluated on the basis of aggregated data. But any conclusions on the overall economic and public health impact, if obtained solely on the basis of aggregated data, are distorted because of the introduction of fixed drug budgets and the effects of the reunification of Germany (among other confounders) that both took place in the beginning of the 1990s, more or less simultaneously with RBP (Schneeweiss et al. 1998).

Sometimes one has to balance the advantage of using individual longitudinal data – without having a control group – against the advantage of having a control

group at the price of rather limited capacities of analysis based on aggregated data. For example, in a study of the effect of premium rebate to reward low utilization of services for beneficiaries of one statutory sickness fund in Germany, the effect on expenditure mainly was analyzed using a long time series of aggregated data together with a control series. In a second step, this analysis was combined with an examination of the effects on non-monetary measures of utilization based on short time series of beneficiary-related data that were primary collected by the sickness fund in order to support administration of premium rebate (Schäfer and Nolde-Gallasch 1999).

24.5 Process of Health Care: Effectiveness, Appropriateness, and Quality

Research on the process of health care considers questions like "Which services are provided in which quantity, by whom, where, and how (Schwartz and Busse 2003)?" The production of health care is a complex result of financing arrangements and of demand- and supply-side factors. The interaction of these different factors is not well understood. An interest in investigating these questions arouse after substantial and unexplained variation in procedures, and hospital admissions were observed between similar hospitals. Some examples for these variations have already been presented in Sect. 24.3.2, for example, the Boston-New Haven Study (Wennberg et al. 1987, 1989). These studies demonstrated the importance of supply-side factors on utilization patterns and frequency, if patient-related factors are controlled for, such as age, sex, case mix, and socioeconomic status. The supply side of a region is primarily described by the density of physicians, hospital beds, and the availability of medical technologies. However, most studies do not analyze the effects of these provider or supply-side characteristics on the health status of the populations concerned (Brook and Lohr 1985). More refined supply-side characteristics determining the use of services comprise provider payment mechanisms, experience and sex of health professionals, organization and equipment of physicians' practices, size and type of hospital, as well as referral patterns between different providers (Schwartz and Busse 2003). "Self-referral" of patients has been identified as an important determinant of small area variation in the use of medical technologies (Childs and Hunter 1972). Self-referral describes the phenomenon of providing expensive medical technology, for example, X-ray examinations, for patients in general practitioners', physicians', or orthopedic surgeons' practices without referring the patient to a radiologist. In comparisons between countries with a comparable standard of health care, the possibility of self-referral for X-ray examinations compared to countries with X-ray examinations exclusively provided by a radiologist increases the overall rate of X-rays by a factor of 4 (Busse 1995). Within a country, differences in examination frequency between doctors with the possibility of self-referral compared to doctors who have to refer patients to a radiologist yield comparable results. In Germany, the rates for X-ray examination for patients with chronic pain were increased by a factor of 2.7, the rates for

abdominal ultrasound for patients presenting with gastrointestinal symptoms by a factor of 3.0 in practices with a possibility of self-referral compared to practices who had to refer their patients to other practices (Busse 1995). Of course this observation is linked to the method of physician remuneration. It is a phenomenon which is primarily observed in countries with fee-for-service remuneration such as the United States and Germany. The effects of the structure of financial incentive systems and resulting overutilization on a system level tend to be underestimated. In Germany, fee-for-service remuneration combined with the possibility of self-referral and the widespread practice of non-radiologists to provide X-ray examinations in their practices resulted in 1,655 X-ray examinations being performed per 1,000 inhabitants in 1997 (Deutsche Röntgengesellschaft 2002) – about twice the rate observed in other European countries. It is estimated that unnecessary X-ray examinations during the last decades now cause around 2,000 incident cases of cancer in the country annually (Berrington de Gonzales and Darby 1994).

The extreme variation in health service provision raises the question whether diagnostic and therapeutic procedures are appropriately used in the process of care. To judge whether a procedure is appropriate, knowledge about the effectiveness of the procedure for certain indications or clinical presentations is required. However, this is not the case for the majority of indication-procedure pairs.

24.5.1 Assessing Effectiveness and Appropriateness of Care

In general, the effectiveness of a health-care professional or service is the degree to which the desired outcomes are achieved (Gray 1997). However, the proposition that an intervention is effective implies that there is only one outcome of care and only one objective in the design of that intervention – which is rarely the case (Gray 1997). In addition to a number of beneficial outcomes of care, such as lower mortality and morbidity, the possibility of harmful effects of care has to be considered. Effectiveness research attempts to answer questions, such as “What is the right thing to do?” or “What care confers significant health benefit for a given clinical situation?” (Scott and Campbell 2002).

Another frequently used concept is that of efficacy, which is the impact of an intervention in the best possible circumstances (Gray 1997). These can be achieved in randomized clinical trials (RCTs). However, the reverse conclusion that RCTs always produce efficacy results is not true, as not all RCTs satisfy high quality standards. The distinction between efficacy and effectiveness is important, as the latter represents the impact of an intervention under routine care conditions. The difference between the two concepts in terms of health status outcomes is illustrated in Fig. 24.2 using the example of complications of radical prostatectomy. Data on effectiveness in the example are derived from a meta-analysis of Medicare routine data; efficacy data are taken from a meta-analysis of RCTs.

An important aspect of both efficacy and effectiveness is that they apply to groups of patients. However, the impact of an intervention on the health status of an individual depends to a large extent on individual factors. To answer the

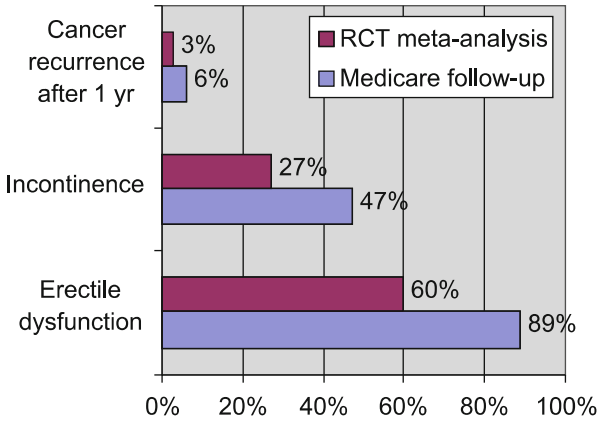

Fig. 24.2 Comparison of effectiveness and efficacy using the example of radical prostatectomy (Adapted from Fowler et al. (1993))

question of whether the most appropriate care was provided given the clinical circumstances is the realm of appropriateness research. An intervention can be considered appropriate, if the expected health benefit exceeds the expected negative consequences by a large enough margin to justify performing the procedure rather than other alternatives (Herrin et al. 1997).

Appropriateness research also addresses the questions of overuse, underuse, or misuse of interventions (Scott and Campbell 2002). We have already discussed the effect of provider remuneration systems and organizational features of a health system on utilization rates. Another major determinant of utilization is clinical judgment. A historical study on clinicians' judgment variation is a study from New York in the 1920s, in which one thousand 11-year-olds had their throats examined (American Child Health Association 1934), cited by Black (1997). In 61% of these children, a tonsillectomy was performed. Of the remaining 39%, the examining doctor thought half would require a tonsillectomy. The half with healthy tonsils were examined by another doctor, who thought that half of them required surgery. The healthy children were again examined by yet another doctor, who declared that half of them required a tonsillectomy, which means that after four examinations only 65 out of 1,000 children would have escaped with their tonsils intact (Black 1997).

During the last 10 years, a number of studies using the RAND/UCLA appropriateness method have been performed. This basically consists in collecting an expert opinion on the appropriateness of an intervention using the Delphi technique. The experts are asked to judge a number of possible indications (that is, case descriptions with clinical information including comorbidity, age, and sex) on whether a specific intervention was appropriate, inappropriate, or even harmful in these cases. In a second step, these judgments are applied to real patient groups in order to determine in how many cases the procedure was appropriate or inappropriate. Most studies using the RAND/UCLA appropriateness method reported appropriate interventions in the range of 50–85% of all interventions performed. Non-US expert groups consistently thought that more procedures were inappropriate compared

to US expert groups. This has to be kept in mind when interpreting results from appropriateness studies. An example is a review of the appropriateness of coronary angiography, upper gastrointestinal endoscopy and carotid endarterectomy performed on 4,564 patients in the USA in 1988. The review which was based on a literature review and on an expert panel consensus concluded that, respectively, only 77%, 76%, and 36% of these procedures had been appropriate (Brook et al. 1990). Inappropriate care is more than a nuisance. It can be harmful to health. For example, in a prospective observational study on congestive heart failure admissions, 7% of admissions were found to be the result of improper medical treatment, including fluid overload, procedures, and misuse of drugs. Hospital mortality for this group of patients was 32% compared to 9% in patients without inappropriate treatment (Rich et al. 1996).

24.5.2 Assessing Quality of Care: Clinical Practice Performance

Measures of clinical practice performance continue to be under constant discussion, particularly when they are published routinely, as is the case, for example, for hospitals in the United States and the United Kingdom. There is broad agreement on the dominant paradigm, established by Donabedian (1980), of measuring quality of clinical care in terms of *structure*, *process*, and *outcome*, but each category has advantages and disadvantages which must be assessed in relation to the type and the speciality of the service (e.g., inpatient vs. outpatient care or surgery vs. drug therapy), the condition being treated (e.g., diabetes, hypertension, acute myocardial infarction, or mental disorders), the case mix of the patients, the role of the comparative information, and a variety of other context variables (Shojania et al. 2001).

The type of assessment of clinical practice performance heavily depends on the perspective on quality. Blumenthal (1996) distinguishes four main perspectives on quality: the health-care professional perspective, the perspective of health-care plans and organizations, the purchaser perspective, and the patient perspective. Health-care professionals tend to emphasize technical excellence and the characteristics of interaction between patient and professional (Donabedian 1988). Health-care plans and organizations place greater emphasis on the general health of the enrolled population and on the function of the organization (Leape 1994). Purchasers, of course, additionally incorporate the price and the effectiveness of the delivery of care. Taking into account the preferences and values of patients leads to a definition of quality that emphasizes outcomes such as functional status, morbidity, mortality, or quality of life and encompasses satisfaction with care (Petitti and Amster 1998).

24.5.2.1 Indicators of Structural Quality

Structural measures characterize the resources in the health system. They describe the setting in which care occurs and the capacity of that setting to produce quality (Donabedian 1980; Brook et al. 1996). Quality assurance programs and organizations such as the Joint Commission on the Accreditation of Health

Care Organizations (JCAHO) and the National Committee on Quality Assurance (NCQA) in the USA or the associations of statutory health insurance physicians in Germany rely on structural measures (as listed below) to infer quality and confer accreditation on this basis.

For providers, structural measures include professional characteristics like speciality or board certifications, etc. For hospitals, they include ownership, number of beds, teaching status, licensure status, availability of sophisticated technologies, qualification of personnel, and other organizational factors for inpatient care (e.g., staff-to-patient ratio, closed intensive care units, dedicated stroke units, or the presence of a clinical information system). One frequently used structural measure of quality is patient volume (Shojania et al. 2001). The growing use of this indicator reflects an extensive literature, which documents superior outcome for hospitals and physicians with higher patient volumes for certain indications and procedures (e.g., Luft et al. 1979; Hannan et al. 1989; Phibbs et al. 1996; Thiemann et al. 1999).

When using structural indicators to measure quality of care, the implicit assumption is that structure affects outcome. This is certainly true for the compliance with minimum standards of structure (e.g., rules for hygiene in operating rooms). But on higher levels of structural quality, the link between structure and outcome is less clear (Shojania et al. 2001). For example, specialist care as a quality measure not always results in better outcomes. This is demonstrated by the findings that even cardiologists fail to provide proven therapies to many eligible patients with acute myocardial infarction (Brand et al. 1995). These findings promote the case to measure the processes of health-care delivery directly instead.

24.5.2.2 Indicators of Process Quality

Process indicators permit a glimpse into the inside of the care-delivering units, allowing measurement of the care patients actually receive. They measure the net effect of physicians' clinical decision making. Clinical choices about the use of surgery, medication or diagnostic tests, admission to a hospital, and length of stay account for a large proportion of the costs of services and of outcomes experienced by the patients. Sometimes generic process measures are used (e.g., number of prescriptions, average length of stay, or day case surgery rate). But mostly they are specific to specialities and certain conditions (e.g., antibiotics within 8 h for patients with community-acquired pneumonia, prophylaxis for venous thromboembolism, or beta-blockers for patients with acute myocardial infarction).

Process measures can be reported for individual physicians, groups of practitioners, for hospitals, hospital units, or hospital trusts, or for the entire system of care. They are favored by providers to indicate quality because they are directly related to what providers do. Frequently they are derived from evidence-based clinical guidelines and facilitate individual physician quality improvement. If proven diagnostic and therapeutic strategies are monitored, quality problems can be detected long before demonstrable outcome differences occur (Brook et al. 1996).

Even so there are some arguments against process-based measurement of the quality of care. First, process measures are not necessarily good predictors of outcome, and allocating resources to processes which do not affect outcomes

may increase cost without producing any improvement in health (Ellwood 1988). Moreover, collecting process data may be a comparatively elaborate procedure. Finally, it may not be possible to achieve consensus on the recommended process for many clinical problems (Petitti and Amster 1998).

24.5.2.3 Indicators of Outcome Quality and Adjustment for Case Mix

The quality-relevant health outcomes have been described as the "five Ds" – death, disease, disability, discomfort, and dissatisfaction (Elinson 1987), or, more positively turned, when measuring quality of care health outcomes could be summarized as survival, states of physiologic, physical, and emotional health, and patient satisfaction (Lohr et al. 1988). Broader definitions of outcomes include psychosocial functioning, quality of life, resource utilization, and costs of care (Iezzoni 1994).

The use of outcome measures to assess the quality of clinical performance has been criticized for several reasons. First, even for common conditions, it may take years to detect differences in outcomes between groups of patients (Palmer 1997). Moreover, such differences may not be under the control of providers but reflect, among others, patient factors, variations in admission practices, or chance rather than differences in quality of care (Shojania et al. 2001). Many outcomes (e.g., mortality) are rare, and comparisons of quality based on such outcomes often have low statistical power (Brook et al. 1996).

Considerable concern is related to perverse incentives for "upcoding" and "gaming" (McGlynn 1998), whereby gaming means a change of treatment to more expensive forms which are frequently more stressful for the patient and result in a reduced quality of care (e.g., a surgical procedure instead of a drug prescription, an inappropriate hospitalization, or a short hospital admission for a marginal diagnosis).

Incentives for gaming may arise from the criteria used to define target patient populations. For example, restricting inpatient mortality to deaths that literally occur in the hospital allows hospitals to lower their mortality rates simply by discharging patients to die at home or in other institutions (Jencks et al. 1988). Additionally, the incentive for physicians or hospitals to avoid caring for sicker patients remains a substantial concern for outcome-based performance measurement (Hofer et al. 1999). Proliferation of diagnoses related to comorbidity or coding of diagnoses related to severity of illness (upcoding) was observed after the introduction of the prospective payment system for HMO-enrolled beneficiaries of Medicare (Keeler et al. 1990).

The most important concern of research related to outcome-based quality measures focused on the development of case-mix adjustment models for hospital mortality rates. Case-mix adjustment and risk adjustment are based on similar methods, but they use different data sources: Case-mix indices are based on medical records from hospitals or physicians, while risk adjustment is based on administrative data, for example, from health insurances. Models that have originally been designed to predict financial rather than clinical outcomes (see Sect. 24.4.1.3) did not perform sufficiently well in this context because hospital data

differ significantly in structure from administrative data. Progress has been made in adopting models to identify the case mix of a group of patients by focusing on specific subgroups of patients instead of overall hospital mortality and using clinical rather than administrative data (Iezzoni 1994).

24.5.3 Examples for Performance Assessment

24.5.3.1 Comparison of HMOs Based on a Performance Indicator System

The federal Centers for Medicare and Medicaid Services (CMS), formerly known as the Health Care Financing Administration (HCA), and the private sector in the USA have supported the development of several performance indicator systems in order to compare the quality of care delivered by HMOs. Perhaps the most popular system, the Health Employer Data and Information Set (HEDIS), was introduced in 1993 and was revised in 1995 and again in 1997. It can be considered as the model for many other performance measurement efforts (Petitti and Amster 1998).

HEDIS was designed by the National Committee for Quality Assurance (NCQA) to evaluate several aspects of health plan performance including clinical quality of care, access to care, satisfaction with care, utilization of services, and the financial performance of the HMO. The clinical quality-of-care indicators included in HEDIS were chosen to address aspects of the care process for which there was strong evidence in the literature to support the relationship between medical care process and desired outcomes (Petitti and Amster 1998). These included indicators for low-birth-weight babies, childhood immunization status, breast cancer screening, eye exams for people with diabetes, and beta-blocker treatment after heart attack.

24.5.3.2 Hospital Ranking

In NHS-type countries, the assessment of the quality of clinical performance focuses on the health of the general population and the function of the health-care system with a special concern on inpatient care. For example, in the United Kingdom, in order to rank hospitals, the publication of clinical indicators in the form of so-called league tables has a long history. As far back as 1983, a set of performance indicators was published covering five areas, one of which was clinical activity. Since then, the set of indicators has been revised several times (British Medical Association 2000). Currently, the published UK league tables are compiled by the Dr Foster organization (separately for England, Wales, Scotland, and North Ireland). They are based on the Department of Health's Hospital Episode Statistics (HES) data and data collected through questionnaires. The indicators fall into five broad categories: standardized mortality rates, waiting times and volumes, staff-to-bed ratios, and – for England only – patient and staff satisfaction and other rating-based scales (clean hospital, good food, etc.). The mortality rates are standardized for age, sex, length of stay, and type of admission (elective or emergency admission). SMRs are calculated for each of 80 ICD9 three-digit primary diagnoses (accounting for 80% of all in-hospital deaths) cited in the final episode of care (Dr Foster 2004a, b). The league tables are criticized for several reasons. The first concern refers to an

insufficient control for case mix relating to severity, comorbidity, deprivation, and the availability of places for people to be discharged to nursing homes or hospices. The second refers to the use of HES data, which are based on finished consultant episodes (the NHS's measure of hospital activity), whereas no conversion to hospital spells is provided (HESs are not designed to collect detailed clinical data). Third, the primary diagnosis has been questioned, as diagnostic criteria change. Finally, the focus on inpatient mortality is considered as a shortfall because an increasing proportion of deaths occur outside the hospital (Jacobson et al. 2003). Dr Foster continued to publish the league tables regardless of this critique, and in the Editors' letter of the 2009 report, you can find the following statement: "Over the years, the report has remained a constant as an independent, authoritative guide to hospital care written for the patient, the politician, the civil servant, the manager and the clinician." (Bedford and Kafetz 2009).

In the United States, an annual index of hospital quality ("America's Best Hospitals") is published by *U.S. News & World Report*. This hospital ranking methodology was devised in 1993 by the statistics and methodology department of the National Organization for Research (NORC) at the University of Chicago. The ranking is based on reputation, mortality, and other factors. The reputational scores of a hospital are based on a survey. The index is designed to be used by patients who are looking for the best hospital to treat their health problems. Since 2005, the annual rankings of "America's Best Hospitals" are produced by the Social, Statistical, and Environmental Sciences Division of RTI International (North Carolina).[9] RTI produces hospital rankings with components representing three key aspects of care: structure, process, and outcome. These components are combined to give an overall score for each hospital in twelve medical specialties. For four additional specialties, scores were based on original survey data alone. The mortality score (outcome) is adjusted for case severity. The severity adjustments were derived using the All Patient Refined Diagnosis Related Group (APR-DRG) method designed by 3M Health Information Systems. The APR-DRG adjusts expected deaths for severity of illness by means of principle diagnosis and categories of secondary diagnoses (RTI 2009).

Beginning in 2007, U.S. News & World Report also began publishing separate annual rankings of "America's Best Children's Hospitals." Like Best Hospitals, the Best Children's Hospitals rankings reflect the interrelationship among structure, process, and outcomes. Most structure and outcome data were obtained directly from children's hospitals using the Pediatric Hospital Survey data submission form which is hosted by RTI. In 2010, children's hospitals were evaluated in ten pediatric medical specialties (RTI 2010).

In Germany, there is no published ranking of hospitals except for some studies of limited impact. A methodology of hospital ranking based on routine data of sickness funds and patient questionnaires with respect to total hip replacement was published

[9]This trade name has ist roots in the Research Triangle Institute (RTI), which was established by the universities located in the Triangle's three cities Raleigh, Durham, and Chapel Hill in North Carolina.

(Schäfer et al. 2007; Bitzer et al. 2007) but until now not fully established for public use. Nevertheless, there are some websites which deliver information on quality of hospital performance (structure, process, and outcome). This information is based on data collected by questionnaires from the insured of statutory sickness funds, on data from the quality reports of the hospitals, established by law in 2004, and – for selected diagnoses and procedures (tracer) – on routine inpatient care data of one large statutory sickness fund. For the latter (risk adjusted) approach (cf. Heller and Günster 2008), the hospitals are rated in the following categories: "above average," "average," and "below average."

24.5.3.3 Physician Profiling

The Physician Payment Review Commission of the American Medical Association (AMA) defines physician profiling as "an analytical tool that uses epidemiological methods to compare practice patterns of providers on the dimensions of cost, service use, or quality (process and outcome) of care." Profiles can be developed for an individual physician, a group of physicians, or physicians within a hospital or managed care plan. They can be broken down by geographical area, speciality, type of practice, or other characteristics. Profiling can focus on many different types of outcome or resource measures. Those resources may be defined globally (e.g., overall charges/costs for the care of a person or group of persons) or they may represent certain subcategories of services (e.g., laboratory, x-ray, physician services, or pharmaceuticals). Profiling is usually applied to compare resources used by cohorts of patients to get a sense of whether their providers do or do not practice efficiently. Even when profiles are not used to modify payment, they may be used to select or reject providers or to determine appropriate patient caseloads for salaried practitioners (Tucker et al. 2002).

The core element of any profiling methodology is risk adjustment by calculating an SMR-like figure, that is, a ratio of observed to expected values of the considered measure. The expected value is adjusted with respect to age, sex of the patients, and to the diagnostic groups that were assigned by the used grouper. Most of the sellers of common models of diagnosis-based risk adjustment (e.g., ACG and DCG) offer the use of their predictive models for profiling.

Profiling may serve as a tool to feed information on care back to the physicians. Also, managed care organizations as a whole have had considerable experience with profiling in order to monitor plan activity. For example, profiling reports, adjusted for case-mix, can be used to distribute bonus or set aside funds that are marked to recognize how well resource management goals are achieved among managed care providers (Tucker et al. 2002). A review of profiling in practice is given by Sutton (2001).

In 1996, the Commonwealth of Massachusetts became one of the first states to implement a comprehensive physician profiling program available to consumers over the Internet. Many other states have adopted similar systems since. In the beginning of the year 2001, physician profiles were available in 30 states, with legislation pending in eight others (Sutton 2001).

In Germany, profiling of physicians has been based on crude measures. Up to now, they have been compared with the average of the regional group of physicians belonging to the same speciality, but a medium-term change to the risk-adjusted profiling system, mandated by law, is scheduled.

24.6 Outcomes of Health Care

24.6.1 Assessing Output and Outcomes of Care

A frequently cited definition of outcomes was given by Donabedian (1985): "Outcomes are those changes, either favorable or adverse, in the actual or potential health status of persons, groups or communities that can be attributed to prior or concurrent care."

The most conventional method of measuring the health status of populations is by means of vital statistics, including statistics of birth and death. Disease-specific incidence rates, cause-specific mortality rates, or other population-based indicators are extensively used to assess the health status of communities, counties, or health systems in general (see Sect. 24.6.3). For example, the Centers for Disease Control (CDC) established a set of 18 population-based health status indicators in 1991 for use at all administrative levels in the United States (Freedman et al. 1991).

Vital statistics may be considered as the key feature of outcome research to study health care and the effect of intervention on a broad range of outcomes, both humanistic and clinical (Petitti 1998a). As population-based measures of health and methods of adjustment are dealt with in chapters ▸Rates, Risks, Measures of Association and Impact and ▸Confounding and Interaction of this handbook and in several sections of this chapter, in the following, we focus on further approaches to measure health status used in outcome research, including patient-based outcomes measurement, adjusted life expectancy, and patient satisfaction. A common feature of most of these outcome measures is that data are collected by questionnaires directly from patients, residents, employees, insured, or HMO-enrolled beneficiaries.

24.6.1.1 Patient-Based Measures of Health Status

Clinicians can make use of a variety of measures which are disease-specific, system- or organ-specific, function-specific (such as instruments that examine sleep or sexual function), or problem-specific (such as back pain) to explore the full range of patients' experience. Disease-specific health status measures have been developed for nearly all chronic conditions, including, for example, asthma, cancer sites, cardiovascular diseases, diabetes, rheumatoid arthritis, prostate disease, epilepsy, hypertension, pneumonia, and migraine (Guyatt et al. 1995). But if there is interest to go beyond the specific illness and to compare the impact of treatments on health-related quality of life (HRQL) across diseases or conditions, one will require a more comprehensive assessment. None of the disease-specific, system- or

organ-specific, function-specific, or problem-specific measures are adequate for comparisons across conditions. These comparisons require generic measures designed for administration to people with any underlying health problem (or no problem at all) that cover all relevant areas of HRQL (Guyatt et al. 1995).

Generic health-status questionnaires are usually designed to establish separate scales including physical, mental, and social health, as suggested by the well-known definition of health by the WHO (1947). There are numerous generic health-status measures – for a review and description, see, e.g., Spilker (1995) or McDowell and Newell (1996). Three of these are very popular and have become standard in the health status field: The 36-Item Short Form Questionnaire (SF-36) (Ware and Sherbourne 1992), the Sickness Impact Profile (SIP) (Bergner et al. 1981), and the Quality of Well-Being Scale (QWB) (Kaplan and Anderson 1988). The psychometric properties of these instruments are sufficiently tested, and the reliability is considered high (Petitti 1998a).

In particular, the SF-36 (a shortened version of a battery of 149 health status questions) is one of the most widely accepted, extensively translated, and tested instruments around the world (Tseng et al. 2003). It satisfies rigorous psychometric criteria for validity and internal consistency. Clinical validity was shown by the distinctive profiles generated for each condition, each of which differed from that in the general population in a predictable manner. Furthermore, SF-36 scores were lower in referred patients than in patients not referred and were closely related to general practitioners' perceptions of severity (Garratt et al. 1993).

The SF-36 was designed for use in clinical practice and research, health policy evaluations, and general population surveys. It includes one multi-item scale that assesses eight health concepts: (1) limitations in physical activities because of health problems; (2) limitations in social activities because of physical or emotional problems; (3) limitations in usual role activities because of physical health problems; (4) bodily pain; (5) general mental health (psychological distress and well-being); (6) limitations in usual role activities because of emotional problems; (7) vitality (energy and fatigue); and (8) general health perceptions. See also the measurement concept in Fig. 24.3 and an excerpt of the questionnaire in Fig. 24.4. The survey was constructed for self-administration by persons 14 years of age and older and for administration by a trained interviewer in person or by telephone (Ware and Sherbourne 1992).

In the late 1980s, a European group of researchers started to develop a generic health-status measure – the European Quality of Life Scale (EQ-5D) – simultaneously in several European languages (EuroQol Group 1990; Brooks 1996). The EuroQol Group consisted originally of a network of international multidisciplinary researchers from Europe but nowadays includes members from Canada, Japan, New Zealand, Singapore, South Africa, and the USA. The EQ-5D self-report questionnaire comprises five dimensions of health (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) rated on three levels (no problems, some/moderate problems, extreme problems). A unique EQ-5D health state is defined by combination of these dimensions. EQ-5D is a public domain instrument (http://www.euroqol.org/index.htm).

Fig. 24.3 The SF-36 measurement concept (Source: SF-36 Psychometric Considerations (http://www.sf-36.org/tools/sf36.shtml))

24.6.1.2 Adjusted Life Expectancy

Life expectancy, even without any adjustments, is already a rather complex measure. It is defined as the average future lifetime of a person at birth and is calculated from a current life table (the key tool of actuaries for some 200 years). Consider a large group, or "cohort," of persons, who were born on the same day. If an actuary could follow the cohort from birth until death, he or she could record the number of individuals alive at each birthday – age x, say – and the number dying during

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks...

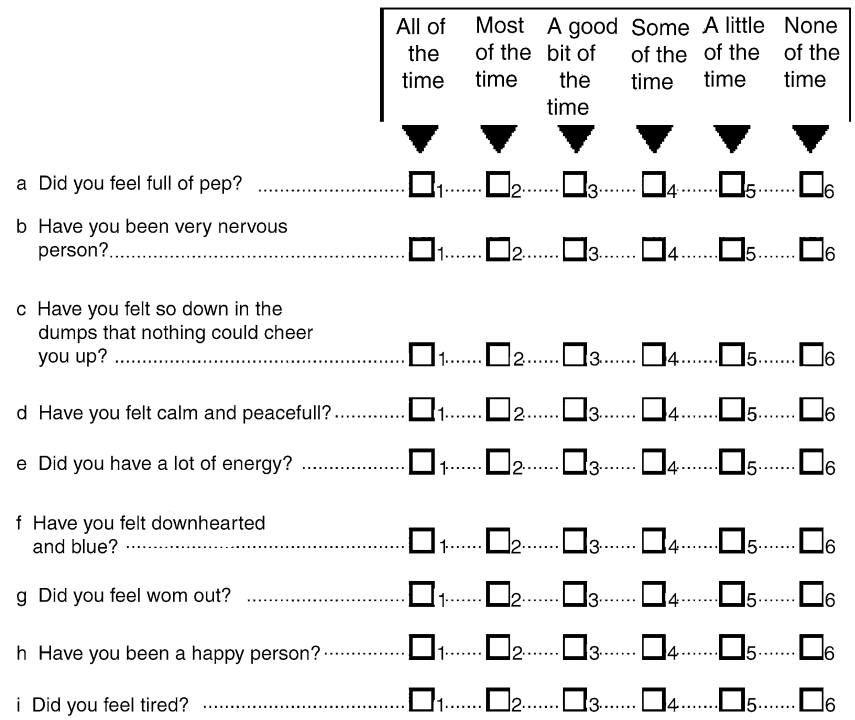

Fig. 24.4 Excerpt from the SF-36 questionnaire: Item 9 (Source: SF-36 Health Survey Scoring Demonstration (http://www.sf-36.org/demos/SF-36.html))

the following year. The ratio of these is the probability of dying at age x, usually denoted by $q(x)$. It turns out that once the $q(x)$'s are all known, the life table is completely determined. In practice, such "cohort life tables" are rarely used, in part because individuals would have to be followed for up to 100 years, and the resulting life table would reflect historical conditions that may no longer have relevance. Instead, one generally works with a period, or current, life table. This summarizes the mortality experience of persons of all ages in a short period, typically 1 year or 3 years. More precisely, the death probabilities $q(x)$ for every age x are computed for that short period, often using census information gathered at regular intervals (e.g., every 10 years in the US). These $q(x)$'s are then applied to a hypothetical cohort of 100,000 people over their life span to create a current life table (Strauss and Shavelle 2000).

Several approaches have been developed to adjust life expectancy for aspects of health-related quality of life (Drummond et al. 1997). Most often used are the concepts of quality-adjusted life years (QALY) on the one hand and the concept of disability-adjusted life years (DALY) on the other hand (cf. chapter ▸Descriptive Studies of this handbook).

A quality-adjusted life year is a measure that assigns a (utility) value, often called Q, between 0 and 1 to each health state of a year with 0 representing death and 1 representing perfect health. The Q factors are then multiplied with the time spent in the corresponding health states, and these weighted times finally are summed up to achieve the QALY.

Three methods are used alternatively to establish a set of consistent Q values, all derived from consumer choice theory, which describes how consumers decide what to buy on the basis of two fundamental elements: their budget constraints and their preferences. Consumer preferences for different consumables are often represented by the concept of "utility" (Torrance et al. 1972; Torrance 1987; Mankiw 1998). The techniques proposed to measure the utility of specific health states on a linear scale were the Von Neumann-Morgenstern "standard gamble," the "time trade-off" method, and direct scaling techniques (e.g., category rating). These were claimed to produce equivalent and reliable results, but the time trade-off is easier to administer than each of the other two techniques (O'Connor 1993). The results of a simultaneous test of all three methods were that subjects found the time trade-off task the easiest, the standard gamble slightly more difficult (but probably impossible without some props), and the direct scaling task the most difficult. Only the time trade-off task was considered to be capable of being executed without a well-trained interviewer (Torrance 1976; O'Connor 1993). Nevertheless, direct scaling methods are commonly used to derive preferences (Petitti 1998a), probably without observing the necessary methodological diligence.

In a standard gamble, the rater (that is, the person to establish the utilities) must choose between two alternatives. One alternative has a certain outcome (that is, the health state to be rated) and the other involves a gamble with two possible outcomes: the best health state (usually complete health), which is described as occurring with a probability, p, or an alternative state, the worst state (usually death) which is described as occurring with probability 1-p. The probability p is varied until the rater is indifferent to the alternative which is certain and the gamble that may bring the better health state. The time trade-off task also entails a choice between two alternatives, but neither is a gamble. Each is a different health state but for differing periods of time. The rater is asked to value a choice of being in a less desirable health state for a longer time followed by death compared with being in a more desirable state for shorter period of time followed by death. The time in the less desirable health state then is decreased to the point of indifference. In category rating, raters sort the health states into a specified number of categories, and equal changes in preference between adjacent categories are assumed to exist (Petitti 1998a).

QALYs have been widely criticized on ethical, conceptual, operational, and methodological grounds. To begin with the last ones, Prieto and Sacristán (2003)

have recently pointed to a considerable problem, which results from the numerical nature of its constituent parts. The appropriateness of the QALY arithmetical operation is compromised by the essence of the utility scale: while life years are expressed in a ratio scale with a true zero, the utility is an interval scale where 0 is an arbitrary value for death. In order to be able to obtain coherent results, both scales would have to be expressed in the same units of measurement. The different nature of these two factors jeopardizes the meaning and interpretation of QALYs. By a simple general linear transformation of the utility scale, the authors demonstrate that the results of the multiplication are not invariant and offer a mathematically solution to these limitations through an alternative calculation of QALYs by means of operations with complex numbers so that the new QALYs have a real part (length of life) and an imaginary part (utility). The revisited formulation of the QALYs provides a less dramatic adjustment of years of life than that implied by the multiplicative model. The maximum penalization represented by living in a suboptimal state of health is capped at 30% of the total time lived in that state, in contrast to the case of the multiplicative model, where the penalization can reach 100% (Prieto and Sacristán 2003).

Ethical concerns arise when QALYs are used in cost-effectiveness or cost-utility analysis for evaluation of alternative health policies, treatment programs, or setting of priorities. A simple ratio cost/QALY is commonly calculated in this type of analysis in order to compare cost-effectiveness of treatments, intervention and programs, etc. But it has been pointed out, among other arguments, that investing in the interventions that have the lowest-cost per QALY ignores the principle of equity (Drummond 1987). In addition, QALYs share a problem of life expectancy as a measure of outcome: they discriminate against the aged and the disabled because these groups of persons have fewer life years to gain from an intervention (Harris 1987).

The main other type of commonly used summary measure which combines information on mortality and morbidity is the disability-adjusted life year. The DALY is the best known example of a "health gap" summary measure, which quantifies the gap between a population's actual health and a defined goal used to quantify the burden of disease in a country, region, or on the global level (Murray and Lopez 1996). However, DALYs share most of the methodological and ethical difficulties with QALYs, such as utility-weighting and discounting health benefits. The discrimination of elderly people is even more pronounced than with QALYs as an additional age-weighting is performed when constructing DALYs, whereby years lost during the productive phase of life get a higher weight than years lost in childhood or at a more advanced age (Gericke and Busse 2003). Related concepts are disability-free life years (Sullivan 1971) and healthy life expectancy (Robine and Ritchie 1991) which may be based on surveys. DALYs can be calculated exclusively based on life tables from census data and cross-sectional data from official disability statistics (if necessary, on a sample base). The so-called Sullivan method to adjust the conventional life table for disability consists of applying disability rates calculated from cross-sectional data to the person-years of the conventional life table. This calculation results into new estimates of the

person-years lived in disability, and the complement of the later, the person-years lived free of disability, the DALYs (Guend et al. 2002).

A related measure is disability-adjusted life expectancy (DALE) used by WHO in a controversial report to display the burden of disease by cause, sex, and mortality stratum in WHO regions (WHO 2000). The disability rates used in the WHO calculations relied on subjective and expert assessment and not on empirical data due to data limitations in many nations.

24.6.1.3 Patient Satisfaction

Interest in measuring satisfaction with health care has grown considerably in recent years around the world, and there is a large and expanding literature in this field. Patient satisfaction and its measurement are undoubtedly important issues for public policy analysts, health-care managers, practitioners, and users. Nevertheless, measurement of satisfaction often lacks a clear definition. In particular, it is not always well understood by the people who measure it that satisfaction is a relative concept which can be measured only against individuals' expectations, needs, or desires (Wüthrich-Schneider 2000; Crow et al. 2002). Despite problems with establishing a tangible definition of "satisfaction" and difficulties with its measurement (among other things those which are predicted by the well-known theory of cognitive dissonance, cf. Festinger 1957), the concept continues to be widely used. However, in many instances, when investigators claim to be measuring satisfaction, more general evaluations of health-care services are being undertaken that tend to result in high levels of satisfaction being recorded (Crow et al. 2002).

Historically patient satisfaction surveys have focused on inpatient health-care services, but in recent years, investigations of patient satisfaction have been carried out in outpatient settings as well. In Germany, for example, a recently developed questionnaire to measure patient satisfaction in generalist and specialist ambulatory medical care comprises 27 single items divided into the four dimensions: "professional competence," "physician-patient interaction," "information," and "organization of the practice." This concept has been tested in a survey of 3,487 patients in 123 physician practices (Gericke et al. 2004). A former international study of patients' priorities with respect to general practice care collected data by postal surveys in UK, Norway, Sweden, Denmark, the Netherlands, Germany, Portugal, and Israel. The study results show that patients in different cultures and health-care systems have many views in common, particularly concerning doctor-patient communication and accessibility of services (Grol et al. 1999).

24.6.2 Assessing Efficiency of Care

In addition to measuring the output of health care in terms of healthy life gained, efficiency is another important dimension in assessing health service output. Unfortunately, the word efficiency is often used inappropriately to describe productivity, that is, relating episodes of care or number of procedures to the inputs or costs (Gray 1997). Efficiency refers to the health system's ability to use whatever

resources it has to maximum effect (Le Grand 1998). Efficiency has three levels: technical, productive, and allocative efficiency. Technical efficiency answers the narrow question of whether the same or a better outcome could be obtained by using less of one type of input (Palmer and Torgerson 1999). It is based on effectiveness. Productive or internal efficiency is achieved when the maximum possible improvement in outcome is obtained from a given level of resource inputs or when costs are minimized to obtain a given level of output (Donaldson and Gerard 1993; Palmer and Torgerson 1999). Prerequisite for productive efficiency is technical efficiency.

Allocative or external efficiency refers to the way resources are divided between alternative uses within the health sector (Barr 1998). It implies productive efficiency (Donaldson and Gerard 1993). The theoretical foundation of allocative efficiency rests on the Pareto criterion: a resource allocation is efficient if it is impossible to move to an alternative allocation which would make some people better off and nobody worse off (Begg et al. 1997b). Among other conceptual difficulties, strict adherence to this principle would preclude changes that would make many people much better off at the expense of a few made slightly worse off (Palmer and Torgerson 1999). Therefore, an operational utilitarian decision rule is often used instead: allocative efficiency is achieved when resource allocation maximizes social welfare (Palmer and Torgerson 1999). Cost-effectiveness studies as a tool to put the concept of operational efficiency in health care into practice have already been summarized in Sect. 24.4.1. Cost-benefit studies can address questions of allocative efficiency comparing interventions between different sectors, as output of care is measured in monetary units. As this is politically and ethically difficult to accept for many non-economists, cost-benefit analyses of health interventions are seldom performed.

24.6.3 Assessing the Outcome of Health Systems

In principle, the same methods are used to assess the outcome of health systems which are used to assess the outcome of health services within a country. However, problems with data quality, definitions, and comparability across different cultures make comparisons between different health systems more difficult than health service research limited to a particular country (Schwartz and Busse 2003). As decision makers in countries of all levels of development are faced with common problems as they struggle to make appropriate choices to improve the performance of their health systems, the interest of politicians and scientists in comparative health system research has grown rapidly during the last two decades. A common goal of researchers is to provide policy decision makers and managers with the best available evidence in order to inform policy decision making. In analogy to evidence-based medicine, this movement has been termed evidence-based health policy or evidence-based health care. However, the evidence-base on how to improve the performance of health systems is still weak (Murray and Evans 2003).

24.6.3.1 Cross-Sectional Comparisons

Two methodological approaches are commonly used in comparative health system research: a cross-sectional approach comparing a number of parameters at a particular point in time and a longitudinal approach comparing the development of parameters over a defined time period. To illustrate the advantages and disadvantages of both approaches, we will focus here on two examples. The first is a summary of the approach taken by the World Health Organization (WHO) to assess health system performance on a global scale. In 1998, WHO embarked on a project to assess the health system performance of its member states, culminating in the World Health Report 2000, in which countries' health systems were ranked according to their performance. Health system performance was measured according to the level and distribution of population health, responsiveness, and fairness in financing (World Health Organization 2000; Murray and Evans 2003). Although the provision of comparative data on health system characteristics is recognized as important in improving health-care systems, the report has elicited heavy criticism, summarized by Gravelle et al. (2003). These included the purpose of the exercise (Williams 2001), the definition of some of the performance measures (Braveman et al. 2001), the quality of data (McKee 2001; Williams 2001), and mixed messages (Navarro 2000). Gravelle et al. (2003) furthermore demonstrated that the efficiency rankings and estimates of the magnitude of inefficiency in countries were not robust when compared with other, no less reasonable, methodological choices concerning the econometric methods used. The final rankings for a number of EU countries and ranking results concerning patient satisfaction with health systems are illustrated in Table 24.4 and compared to a number of other parameters which are commonly used to measure the input, process, and outcome of a health system.

It can be noted that parameters differ widely between countries at a similar level of national income and development. Some factors show a close correlation, for example, the health score with patient satisfaction or hospital bed provision with hospital utilization. On the other hand, satisfaction with the health system does not correlate at all with the overall WHO ranking of the health system performance. This results in contradictory results for countries like Denmark and Finland on the one hand, and Spain on the other (Schwartz and Busse 2003). Table 24.4 illustrates some of the issues surrounding the interpretation of cross-sectional data. Different data sources can vary substantially on the same measure. Such discrepancies – if detected at all – demand a thorough investigation of possible causes. A common cause are differences in the numerator, for example, differences between licensed and practicing doctors or beds in acute care hospitals or in all hospitals. Differences in the denominator are also important. For instance, for the measurement of neonatal mortality, it makes a difference whether all births on the territory of a country are counted or all births of nationals of that country (Schwartz and Busse 2003).

The most important difficulty with cross-sectional comparisons of health systems from a policy perspective is probably that health output measured in terms of reduced mortality and health system performance are correlated in a contradictory way. If a country reacts in an appropriate way to high mortality rates and invests in

Table 24.4 Selected input, process, and outcome parameters for some European countries, around 1997. Data from the OECD Health Dataset 2001, the WHO Health for All database 2003, and the World Health Report 2000 (World Health Organization 2000) (Adapted from Schwartz and Busse (2003))

	Financial input: % of GDP (1998)	Structure: Hospital beds/1,000 population (1997)	Structure: Doctors/ 1,000 population (1997)	Process: Hospital cases/100 pop./year (1996)	Process: Hospital days/capita (1996)	Process: Ambulatory doctor-patient contacts/year (1996)	Outcome: Neonatal mortality/ 1,000 (1998)	Outcome: Satisfaction with health system in % [Ranking within EU] (1998)	Outcome: Overall ranking of health system within EU (1999)
Austria	8.0	9.1	2.9	25.1	2.6	6.3	4.9	72.7 [3]	4
Belgium	8.6	7.3	3.7	20.0	2.2	8.0	5.6	62.8 [7]	11
Denmark	8.3	4.6	3.3	19.8	1.4	5.7 (6.6)[a]	4.7	90.6 [1]	15
Finland	6.9	7.9	3.0	26.9	3.2	4.3	4.1	81.3 [2]	14
France	9.4	8.6	3.0	22.5	2.6	6.5	4.6	65.0 [6]	1
Germany	10.3	9.4	3.4	19.7	2.8	6.5	4.7	57.5 [9]	13
Greece	8.4	5.0	4.1	–	1.2	–	5.7 (6.7)[a]	15.5 [15]	6
Ireland	6.8	–	2.1	15.1	1.1	–	6.2	57.9 [8]	10
Italy	8.2	5.8	5.8	18.5	1.7	–	5.3	20.1 [13]	2
Luxembourg	6.0	8.1	3.0 (2.4)[a]	–	2.8	2.9	5.0	66.6 [5]	7
Netherlands	8.7	11.3 (5.3)[a]	–	11.1	3.6	5.4	5.0	69.8 [4]	8
Portugal	7.7	4.1	3.1	11.4	1.1	3.2	5.9	16.4 [14]	5
Spain	7.0	–	2.9	11.4 (10.0)[a]	1.1	–	5.7 (4.9)[a]	43.1 [12]	3
Sweden	7.9	4.0 (5.2)[a]	3.1	18.1	1.3	2.9	3.5	57.5 [9]	12
United Kingdom	6.8	4.4	1.7	15.0 (23.1)[a]	1.3	6.1	5.7	57.0 [11]	9

[a]More than 10% difference between OECD and WHO datasets

health system infrastructure, mortality would fall as a result assuming effectiveness of the measures taken. This longitudinal result cannot be measured in cross-sectional studies. Therefore cross-sectional comparisons cannot indicate whether a high level of inputs in a particular country has obviated even higher mortality rates and we only see average mortality in this country or whether there truly exists an inefficient input-output relation.

24.6.3.2 Longitudinal Comparisons

The other approach consists in comparing the development of input, process, and output parameters in different health systems in a longitudinal perspective. In the 1980s, time series analyses on "avoidable mortality" marked the first attempts at international longitudinal comparisons (Bunker et al. 1994; Charlton and Velez 1986). A common measure for comparing health systems in a longitudinal way is life expectancy. In Fig. 24.5, the development of life expectancy at birth is depicted for a number of selected European countries compared to the EU average for the time period 1970–2000.

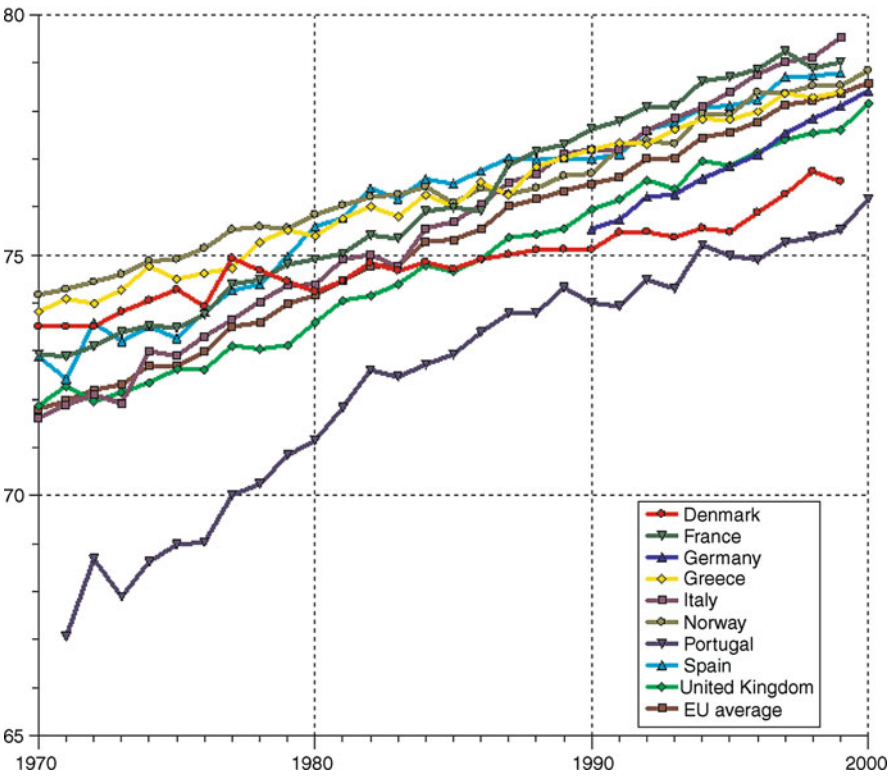

Fig. 24.5 Life expectancy at birth in selected member countries of the European Union 1970–2000. Calculated with data from WHO Health for All database 2003

This is often done although it is well known that life expectancy is influenced by many variables outside the scope of the health system, such as the level of socioeconomic development. However, in general, mortality (on which the calculation of life expectancy is based) is an output measure which is relatively insensitive to common health service endeavours (Schwartz and Busse 2003):

- The overwhelming part of mortality is not amenable to health service activity ("avoidable mortality") but natural mortality.
- In particular for men, a substantial proportion of deaths is due to traffic accidents.
- The commonplace argument that mortality figures do not respond quickly to changes had to be revised after the experience in Russia after the breakdown of the Soviet Union, where life expectancy at birth for males fell by approximately 6 years between 1990 and 1994. Life expectancy in the Eastern part of Germany, however, increased substantially in the 1990s.

Relative changes over time are of particular importance for evaluation and policy decision making. This is illustrated in the development of life expectancy at birth in Denmark and Portugal. Whereas both countries had an average life expectancy at birth of 76 years in the year 2000, Portugal has massively improved on this measure since 1970, up from 67 years. Although life expectancy in Denmark has nominally also increased since 1970, up from 74 years, it has had the smallest relative increase in Western Europe – which is in fact a rather negative development and not an improvement.

24.7 Conclusions

As demonstrated in the examples discussed above, the combination of simple inputs and outputs can be of particular political importance, despite all the methodological difficulties and caveats. The fact that even if life expectancy were a good indicator of health production in the health-care system, the question of why a good result has occurred, that is, examining structure and process, would not have been answered. There is little consensus on how international comparisons of structures and processes should be performed. How inappropriate simplification of health system comparisons can be misleading is demonstrated by the "state versus free market" debate in Germany. Financing of German hospital care on the basis of *per diem* payments has been coined as inefficient, as this payment mechanism creates an incentive for longer hospital stays. Some economists have compared the German system with the US system, where hospital stays are usually shorter, claiming that this was due to payments according to diagnostic-related groups (DRGs). However, they did not consider that at that time, only hospital services for 15% of the population covered under the Medicare scheme were remunerated according to DRGs and that hospital costs per case in the USA were about twice as high as in Germany, "despite" the DRGs. Likewise, the expected rise in ambulatory care costs to compensate for early hospital discharge was not considered (Schwartz and Busse 2003).

International comparisons of health system outcomes along one-dimensional hypotheses have thus to be treated with great caution, in particular, because they are easily misunderstood by policy decision makers (Schwartz and Busse 2003).

References

Abraham S (1986) Analysis of data from a complex sample: the Health Examination Surveys. Am J Clin Nutr 43:839–843

ACRA (1999) A brief history of resource allocation in the NHS, 1948–98. Advisory Committee on Resource Allocation, Department of Health, London

Aday LA, Begley CE, Lairson DR, Slater CH (1998) Evaluating the healthcare system. Effectiveness, efficiency, and equity. Health Administration Press, Chicago

Adler GS (1994) A profile of the Medicare Current Beneficiary Survey. Health Care Financ Rev 15(4):153–163

American Child Health Association (1934) Physical defects: the pathway to correction. ACHA, New York

Armenian HK (1998) Case-control methods. In: Amenian HK, Shapiro S (eds) Epidemiology and health services. Oxford University Press, New York/Oxford, pp 135–155

Ash A, Porell F, Gruenberg L, Sawitz E, Beiser A (1989) Adjusting Medicare capitation payments using prior hospitalization data. Health Care Financ Rev 10(4):17–29

Asthana S, Gibson A, Moon G, Dicker J, Brigham P (2004) The pursuit of equity in NHS resource allocation: should morbidity replace utilisation as the basis for setting health care capitations? Soc Sci Med 58:539–551

Barr N (1998) The economics of the welfare state. Oxford University Press, Oxford

Bedford Z, Kafetz A (2009) Editors' letter. In: The Dr Foster Hospital Guide 2009. Dr Foster Research. London

Begg D, Fischer S, Dornbusch R (1997a) Demand, supply, and the market. In: Economics. McGraw Hill, London, pp 30–43

Begg D, Fischer S, Dornbusch R (1997b) Introduction to welfare economics. In: Economics. McGraw Hill, London, pp 240–259

Bellach B-M, Knopf H, Thefeld W (1998) Der Bundesgesundheitssurvey 1997/98. Gesundheitswesen 60(Suppl 2):59–68

Bennett AC (1978) Improving management performance in health care institutions. American Hospital Association, Chicago

Bergmann E, Kamtsiuris E (1999) Inanspruchnahme medizinischer Leistungen. Gesundheitswesen 61(Spec Issue):138–144

Bergner M, Bobbitt RA, Carter WB, Gilson BS (1981) The sickness impact profile: development and final revision of a health status measure. Med Care 19:787–805

Berrington de Gonzales A, Darby S (1994) Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. Lancet 363:345–351

Bitzer EM, Neusser S, Lorenz C, Dörning H, Schäfer T (2007) Krankenhaus-Rangfolgen nach Ergebnisqualität in der Hüftendoprothetik - Routinedaten mit oder ohne Patientenbefragungen? - Teil 2: Patientenbefragung in Kombination mit Routinedaten. GMS Med Inf Biom Epidemiol 3(1):Doc07

Bitzer EM, Grobe T, Dörning H, Schwartz FW (2010) BARMER GEK Report Krankenhaus 2010. Asgard, St. Augustin

Black N (1997) Health services research: saviour or chimera? Lancet 349:1834–1836

Black N, Langham S, Petticrew M (1995) Coronary revascularisation: why do rates vary geographically in the UK? J Epidemiol Community Health 49:408–412

Blumenthal D (1996) Quality of care: what is it? N Engl J Med 335:891–894

Brand DA, Newcomer LN, Freiburger A, Tian H (1995) Cardiologists' practices compared with practice guidelines: use of beta-blockade after acute myocardial infarction. J Am Coll Cardiol 26:1432–1436

Braveman P, Starfield B, Geiger HJ (2001) World Health Report 2000: how it removes equity from the agenda for public health monitoring and policy. BMJ 323:678–681

British Medical Association (2000) Clinical indicators (League tables) – a discussion paper. British Medical Association, London

Brook RH, Lohr KN (1985) Efficacy, effectiveness, variations, and quality. Boundary-crossing research. Med Care 23:710–722

Brook RH, Ware JE Jr, Rogers WH, Keeler EB, Davies AR, Donald CA, Goldberg GA, Lohr KN, Masthay PC, Newhouse JP (1983) Does free care improve adults' health? Results from a randomized controlled trial. N Engl J Med 309:1426–1434

Brook RH, Park RE, Chassin MR, Solomon DH, Keesey J, Kosecoff J (1990) Predicting the appropriate use of carotid endarterectomy, upper gastrointestinal endoscopy, and coronary angiography. N Engl J Med 323:1173–1177

Brook RH, McGlynn EA, Cleary PD (1996) Quality of health care. Part 2: Measuring quality of care. N Engl J Med 335:966–970

Brooks R (1996) EuroQol: the current state of play. Health Policy 37:53–72

Brückner G (1997) Developing a new system of health care statistics. A major challenge for all participants? Federal Statistical Office, Wiesbaden

Bunker JP, Frazier HS, Mosteller F (1994) Improving health: measuring effects of medical care. Milbank Q 72:225–258

Buring JE, Hennekens CH (1992) The women's health study: summary of the study design. J Mycardial Ischemia 4:27–29

Busse R (1995) Radiologie, Gesundheitsstrukturreform und Gesundheitssystemforschung – Stand, Entwicklungen und Herausforderungen. Akt Radiologie 5:127–130

Busse R, Wismar M (2002) Health target programmes and health care services – any link? A conceptual and comparative study (part 1). Health Policy 59:209–221

Cameron AC, Trivedi PK (1998) Regression analysis of count data. Cambridge University Press, Cambridge/New York/Melbourne

Campell DT, Stanley JC (1966) Experimental and quasi-experimental designs for research. Rand McNally, Chicago

Carr-Hill R (1989) Allocating resources to health care: RAWP (Resources Allocation Working Party) is dead-long live RAWP. Health Policy 13:135

Carr-Hill RA, Sheldon TA, Smith P, Martin S, Peacock S, Hardman G (1994) Allocating resources to health authorities: development of method for small area analysis of use of inpatient services. BMJ 309(6961):1046–1049

Charlton JR, Velez R (1986) Some international comparisons of mortality amenable to medical intervention. BMJ 292:295–301

Childs AW, Hunter ED (1972) Non-medical factors influencing use of diagnostic x-ray by physicians. Med Care 10:323–335

Cochran WG (1968) Sampling techniques. Wiley, New York

Cohen SB (1997) Sample design of the 1996 Medical Expenditure Panel Survey Household Component. MEPS Methodol Rep No 2. AHCPR Pub. No 97–0027, Rockville

Crow R, Gage H, Hampson S, Hart J, Kimber A, Storey L (2002) The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. Health Technol Assess 6:32

Culyer A (1993) Health, health expenditures, and equity. In: VanDoorslaer E (ed) Equity in the finance and delivery of health care. Oxford University Press, Oxford

de Miguel JM (1971) A framework for the study of national health systems. Inquiry 12:10–24

Department of Health (2008) Resource allocation: weighted capitation formula, 6th edn. Department of Health, Leeds

Detsky AS, Naglie IG (1990) A clinician's guide to cost-effectiveness analysis. Ann Intern Med 113:147–154

Deutsche Röntgengesellschaft (2002) Pressemitteilung anlässlich des 83. Deutschen Röntgenkongresses vom 8. - 11. Mai 2002 in Wiesbaden

Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY (1999) Methods for analyzing health care utilization and costs. Ann Rev Public Health 20:125–144

Donabedian A (1980) Explorations in quality assessment and monitoring. Health Administration Press, Ann Arbor, MI

Donabedian A (1985) The methods and findings of quality assessment and monitoring. Health Administration Press, Ann Arbor, MI

Donabedian A (1988) The quality of care: how can it be assessed? JAMA 260:1743–1748

Donaldson C, Gerard K (1993) Economics of health care financing: The visible hand. Macmillan, London

Doubilet P, Weinstein MC, McNeil JB (1986) Use and misuse of the term 'cost-effectiveness' in medicine. N Engl J Med 314:253–256

Dr Foster (ed) (2004a) Hospital guide – methodology. http://www.drfosterhealth.co.uk/hospital-guide/methodology/. Accessed 29 Mar 2011

Dr Foster (ed) (2004b) Consultant guide – methodology. http://www.drfosterhealth.co.uk/consultant-guide/methodology.aspx. Accessed 29 Mar 2011

Drummond MF (1987) Ressource allocation decisions in health care. A role for quality of life assessments? J Chronic Dis 40:605–616

Drummond MF, O'Brien B, Stoddart GL, Torrance GW (1997) Methods for the economic evaluation of health care programmes, 2nd edn. Oxford University Press, Oxford

Dunn DL, Rosenblatt A, Taira DA, Lattimer E, Bertko J, Stoiber T (1996) A comparative analysis of methods of health risk assessment. Society of Actuaries, Schaumburg, IL

Elinson J (1987) Advances in health assessment discussion panel. J Chronic Dis 40(Suppl 1): 83S–91S

Ellis R, Pope G, Iezzoni L, Ayanian J, Bates D, Burstin H, Ash A (1996) Diagnosis-based risk adjustment for Medicare capitation payments. Health Care Financ Rev 17(3):101–128

Ellwood PM (1988) Outcomes management: a technology of patient experience. N Engl J Med 318:1549–1556

Epstein RS, Sherwood LM (1996) From outcomes research to disease management. A guide for the perplexed. Ann Intern Med 124:832–837

Etchason J, Petz L, Keeler E, Calhoun L, Kleinman S, Snider C, Fink A, Brook R (1995) The cost effectiveness of preoperative autologous blood donations. N Engl J Med 332:719–724

EuroQol Group (1990) EuroQol – a new facility for the measurement of health-related quality of life. Health Policy 16:199–208. http://www.euroqol.org/. Accessed 29 Mar 2011

Fein R (1971) On measuring economic benefits of health programs. In: McLachlan G, McKeown T (eds) Medical history and medical care. Oxford University Press, London

Ferris G, Roderick P, Smithies A, George S, Gabbay J, Couper N, Chant A (1998) An epidemiological needs assessment of carotid endarterectomy in an English health region. Is the need being met? BMJ 317:447–451

Festinger L (1957) A theory of cognitive dissonance. Stanford University Press, Stanford, CA

Fowler FJ Jr, Barry MJ, Lu-Yao G, Roman A, Wasson J, Wennberg JE (1993) Patient-reported complications and follow-up treatment after radical prostatectomy. The national Medicare experience: 1988–1990 (updated June 1993). Urology 42:622–629

Freedman MA in collaboration with the CDC Health Status Indicators Consensus Work Group (1991) Health status indicators for the year 2000. Health People 2000 Statistical Notes, vol 1 (no. 1). National Center for Health Statistics, Hyattsville

Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT (1993) The SF36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? BMJ 306: 1440–1444

Gericke C, Busse R (2003) Gesundheitsökonomische Aspekte der Pharmakotherapie älterer Menschen. Arzneimittelforschung Drug Res 53:918–921

Gericke CA, Schiffhorst G, Busse R, Häussler B (2004) Ein valides Instrument zur Messung der Patientenzufriedenheit in ambulanter haus- und fachärztlicher Behandlung: das QUALISKOPE-A. Das Gesundheitswesen 66:723–731

Gillings D, Makuc D, Siegel W (1981) Analysis of interrupted time series mortality trends; an example to evaluate regionalized perinatal care. Am J Public Health 71:38–46

Glaeske G, Schicktanz C (2010) BARMER GEK Arzneimittel-Report 2010. Asgard, St. Augustin

Gold MR, Siegel JE, Rusek KB, Weinstein MC (1996) Cost-effectiveness in health and medicine. Oxford University Press, New York

Gravelle H, Jacobs R, Jones AM, Street A (2003) Comparing the efficiency of national health systems: a sensitivity analysis of the WHO approach. Appl Health Econ Health Policy 2: 141–147

Gray JAM (1997) Assessing the outcomes found. In: Evidence-based healthcare. Churchill Livingstone, New York, pp 103–154

Greene WH (2003) Econometric Analysis, 5th edn. Prentice and Hall, Upper Saddle River

Grimshaw JM, Wilson B, Campbell M, Eccles M, Ramsay C (2001) Epidemiological methods. In: Fulop N, Allen P, Clarke A, Black N (eds) Studying the organisation and delivery of health services: research methods. Routledge, London, pp 56–72

Grobe T, Dörning H, Schwartz FW (2006) GEK-Report ambulant-ärztliche Versorgung 2006. Asgard, St. Augustin

Grobe T, Dörning H, Schwartz FW (2011) BARMER GEK Arztreport 2011. Asgard, St. Augustin

Grol R, Wensing M, Mainz J, Ferreira P, Hearnshaw H, Hjortdahl P, Olesen F, Ribacke M, Spenser T, Szecsenyi J (1999) Patients' priorities with respect to general practice care: an international comparison. Fam Pract 16:4–11

Guend H, Stone-Newsom R, Swallen K, Lasker A, Kindig D (2002) State disability adjusted life expectancy using census disability. University of Wisconsin, Madison

Guyatt GH, Naylor CD, Juniper E, Heyland DK, Jaeschke R, Cook RD for the Evidence Based Medicine Working Group (1995) How to use articles about health-related quality of life measurements. Centre for Health Evidence, Edmonton. http://www.cche.net/usersguides/life.asp. Accessed 28 May 2004

Hannan EL, O'Donnell JF, Kilburn H Jr, Bernard HR, Yazici A (1989) Investigation of the relationship between volume and mortality for surgical procedures performed in New York State hospitals. JAMA 262:503–510

Hansen MH, Hurwitz WN, Madow WG (1953) Sample survey methods and theory, vols I and II. Wiley, New York/London

Harris J (1987) QUALYfying the value of life. J Med Ethics 13:117–123

Heller G, Günster C (2008) Mit Routinedaten Qualität in der Medizin sichern. GGW 8:26–34

Herrin J, Etchason JA, Kahan JP, Brook RH, Ballard DJ (1997) Effect of panel composition on physician ratings of appropriateness of abdominal aortic aneurysm surgery: elucidating differences between multispecialty panel results and specialty society recommendations. Health Policy 42:67–81

Hill AB (1965) The environment and disease: association or causation? Proc R Soc Med 58: 295–300

Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG (1999) The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. JAMA 281:2098–2105

Iezzoni LI (1994) Risk and outcome. In: Iezzoni LI (ed) Risk adjustment for measuring healthcare outcomes. Health Administration Press, Ann Arbor, Michigan, pp 1–28

Ihle P, Köster I, Küpper-Nybelen J, Schubert I (2008) Experiences with a person-related and population-based sickness fund sample (1997–2007) for pharmacoepidemiological and health care utilization research. Abstract of the 24th International Conference on Pharmacoepidemiology & Therapeutic Risk Management. Copenhagen, Denmark, August 17–20, 2008 Pharmacoepidemiol Drug Saf 17(Suppl 1): S242

Ingber MJ (1998) The current state of risk adjustment technology for capitation. J Ambul Care Manag 21(4):1–28

Institute of Medicine (1994) Health services research: opportunities for an expanding field of inquiry – An interim statement. National Academies Press, Washington, DC
Jacobson B, Mindell J, McKee M (2003) Hospital mortality league tables. BMJ 326:777–778
Jencks SF, Williams DK, Kay TL (1988) Assessing hospital-associated deaths from discharge data. The role of length of stay and comorbidities. JAMA 260:2240–2246
Kahn HA, Sempos CT (1989) Statistical methods in epidemiology. Oxford University Press, New York/Oxford
Kaplan RM, Anderson JP (1988) A general health policy model: update and applications. Health Serv Res 23:203–205
Keeler EB, Kahn KL, Draper D, Sherwood MJ, Rubenstein LV, Reinisch EJ, Kosecoff J, Brook (1990) Changes in sickness at admission following the introduction of the prospective payment system. JAMA 264:1962–1968
Kelsey JL, Petitti DB, King AC (1998) Key methodologic concepts and issues. In: Brownson RC, Petitti DB (eds) Applied epidemiology. Oxford University Press, New York, Oxford, pp 35–69
Kendall MG, Stuart A (1958) Advanced theory of statistics, vol 1. Charles Griffin and Co., London
Kindig DA (1997) Different populations, different needs? In: Purchasing population health: paying for results. The University of Michigan Press, Ann Arbor, pp 133–148
Kish L (1965) Survey sampling. Wiley, New York
Kitagawa EM, Hauser PM (1973) Differential mortality in the United States – a study in socioeconomic epidemiology. Harvard University Press, Cambridge
Kjerulff KH, Erickson BA, Langenberg PW (1996) Chronic gynecological conditions reported by U.S. women: finding from the National Health Interview Survey 1984 to 1992. Am J Public Health 86:195–199
Kohn R, White KL (eds) (1976) Health care – an international study. Oxford University Press, Oxford/New York/Toronto
Krishnaiah PR, Rao CR (eds) (1994) Handbook of statistics 6 - Sampling, 2nd edn. Elsevier Science Publishers, Amsterdam
Kronick R, Dreyfus T, Lee L, Zhou Z (1996) Diagnostic risk adjustment for Medicaid: the disability payment system. Health Care Financ Rev 17(3):7–33
Kronick R, Gilmer T, Dreyfus T, Ganiats T (2002) CDPS-Medicare: the chronic illness and disability payment system modified to predict expenditures for Medicare beneficiaries. Final report to CMS. University of California, San Diego
Kurth B-M (2007) The German Health Interview and Examination Survey for Children and Adolescents (KiGGS): an overview of its planning, implementation and results taking into account aspects of quality management. In: KiGGS – Principal publication, methodology and conduct of field work. Bundesgesundheitsbl – Gesundheitsforsch – Gesundheitsschutz 50(5–6):533–546
Lamers LM (1999) Pharmacy costs groups: a risk-adjusted for capitation payments based on the use of prescribed drugs. Med Care 37(8):824–830
Last JM (2001) A dictionary of epidemiology, 4th edn. Oxford University Press, Oxford
Le Grand J (1998) Financing health care. In: Feachem Z, Hensher M, Rose L (eds) Implementing health sector reform in Central Asia. World Bank, Washington, DC, pp 75–85
Leape LL (1994) Error in medicine. JAMA 272:1851–1857
Levy PS, Lemesshow S (1991) Sampling of populations: methods and applications. Wiley, New York
Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13–22
Lipid Research Clinics Program (1984) The lipid research clinics coronary primary prevention trial results I. Reduction in incidence of coronary heart disease. JAMA 251:351–364
Lohr KN, Yordi KD, Their SO (1988) Current issues in quality of care. Health Aff 7:5–18
Luft HS, Bunker JP, Enthoven AC (1979) Should operations be regionalized? The empirical relation between surgical volume and mortality. N Engl J Med 301:1364–1369
Mankiw NG (1998) Principles of economics Harcourt. Brace, Boston

Mays N (1995) Geographical resource allocation in the English National Health Service, 1974–1994: the tension between normative and empirical approaches. Int J Epidemiol 24: 96–102

McCullagh P, Nelder JA (1983) Generalized linear models. Chapman and Hall, London/New York

McDowell I, Newell C (1996) Measuring health: a guide to rating scales and questionnaires, 2nd edn. Oxford University Press, New York/Oxford

McGlynn EA (1998) The outcomes utility index: Will outcomes data tell us what we want to know? Int J Qual Health Care 10:485–490

McKee M (2001) Measuring the efficiency of health systems. The world health report sets the agenda, but there's still a long way to go. BMJ 323:295–296

McPake B, Kumaranayake L, Normand C (2002) The demand for health and health services. In: Health economics. An international perspective. Routledge, London/New York, pp 12–19

Medicare Payment Advisory Commission (1998) Report to the Congress: context for a changing Medicare program. Medicare Payment Advisory Commission, Washington

Meyer N, Fischer R, Weitkunat R, Crispin A, Schotten K, Bellach B-M, Überla K (2002) Evaluation des Gesundheitsmonitorings in Bayern mit computer-assistierten Telefoninterviews (CATI) durch den Vergleich mit dem Bundesgesundheitssurvey 1998 des Robert Koch-Instituts. Gesundheitswesen 64:329–335

Murray CJL, Lopez AD (1996) The global burden of disease. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Harvard University Press, Cambridge, MA

Murray JL, Evans DB (2003) Health systems performance assessment: goals, framework and overview. In: Murray JL, Evans DB (eds) Health systems performance assessment: debates, methods and empiricism. World Health Organization, Geneva, pp 3–20

Navarro V (2000) Assessment of the world health report 2000. Lancet 356:1598–1601

Nelder JA, Wedderburn EWM (1972) Generalized linear models. J R Stat Soc A 135:370–384

Neter J, Wasserman W (1974) Applied linear statistical models. Richard D. Irwin Inc., Homewood, IL

Neuhauser D, Lewicki AM (1975) What do we gain from the sixth stool guaiac? N Engl J Med 293: 226–228

Newhouse JP (1974) A design for a health insurance experiment. Inquiry 11:5–27

Newhouse JP (1986) Rate adjuster for Medicare under capitation. Health Care Financ Rev (Spec No):45–55

Newhouse JP (1993) Free for all? Lessons from the RAND health insurance experiment. Harvard University Press, Cambridge, MA

Newhouse JP, McClellan M (1998) Econometrics in outcomes research: the use of instrumental variables. Ann Rev Public Health 19:17–34

Newhouse JP, Manning WG, Morris CN, Orr LL, Duan N, Keeler EB, Leibowitz A, Marquis KH, Marquis MS, Phelps CE, Brook RH (1981) Some interim results from a controlled trial of cost sharing in health insurance. N Engl J Med 305:1501–1507

Newhouse JP, Manning WG, Keeler EB, Sloss EM (1989) Adjusting capitation rates using objective health measurers and prior utilization. Health Care Financ Rev 10(3):41–54

Newhouse JP, Buntin MB, Chapman JD (1997) Risk adjustment and Medicare: taking a closer look. Health Affairs 16(5):26–43

O'Connor R (1993) Issues in the measurement of health-related quality of life. Working paper 30. NHMRC National Centre for Health Program Evaluation Melbourne, Australia

OECD (1993) System of national accounts. OECD, Paris

OECD (2000) A system of health accounts, Version 1.0. OECD, Paris

Palmer RH (1997) Process-based measures of quality: the need for detailed clinical data in large health care databases. Ann Intern Med 127:733–738

Palmer S, Torgerson DJ (1999) Definitions of efficiency. BMJ 318:1136

Petitti DB (1998a) Epidemiological issues in outcomes research. In: Brownson RC, Petitti DB (eds) Applied epidemiology. Oxford University Press, New York/Oxford, pp 249–275

Petitti DB (1998b) Economic evaluation. In: Brownson RC, Petitti DB (eds) Applied epidemiology. Oxford University Press, New York/Oxford, pp 277–298
Petitti DB, Amster A (1998) Measuring the quality of health care. In: Brownson RC, Petitti DB (eds) Applied epidemiology. Oxford University Press, New York/Oxford, pp 299–321
Petitti DB, Sidney S (1989) Hip fracture in women: incidence, in-hospital mortality, and five-year survival probabilities in members of prepaid health plan. Clin Orthop 246:150–155
Phibbs CS, Bronstein JM, Buxton E, Phibbs RH (1996) The effects of patient volume and level of care at the hospital of birth on neonatal mortality. JAMA 276:1054–1059
Pigeot I, Ahrens W (2008) Establishment of a pharmacoepidemiological database in Germany: methodological potential, scientific value and practical limitations. Pharmacoepidemiol Drug Saf 17:215–223
Pope GC, Ellis RP, Ash AS, Liu C-F, Ayanian JZ, Bates DW, Burstin H, Iezzoni LI, Ingber MJ (2000) Principal inpatient diagnostic cost group model for Medicare risk adjustment. Health Care Financ Rev 21(3):93–118
Pope GC, Kautter J, Ellis RP, Ash AS, Ayanian JZ, Iezzoni LI, Ingber JM, Levy JM, Robst J (2004) Risk adjustment of Medicare capitation payments using the CMS-HCC model. Health Care Financ Rev 25(4):119–141
Prieto L, Sacristán JA (2003) Problems and solutions in calculating quality-adjusted life years (QALYs). Health Qual Life Outcome 1:80. http://www.hqlo.com/content/1/1/80. Accessed 30 July 2013
Rice N, Smith P (1999) Approaches to capitation and risk adjustment in health care: an international survey. ACRA paper 09, Department of Health, London
Rich MW, Shah AS, Vinson JM, Freedland KE, Kuru T, Sperry JC (1996) Iatrogenic congestive heart failure in older adults: clinical course and prognosis. J Am Geriatr Soc 44:638–643
Robert Koch Institute (2008a) German Health Interview and Examination Survey for Adults (DEGS). http://www.rki.de/cln_160/nn_217400/EN/Content/Health_Reporting/HealthlSurveys/Degs/degs_node.html?_nnn=true. Accessed 28 Mar, 2011
Robert Koch Institute (2008b) GEDA: telephone health survey 2008/2009. http://www.rki.de/cln_160/nn_675636/EN/Content/Health_Reporting/HealthlSurveys/Geda/ Geda_ node.html?_ nnn = true. Accessed 28 Mar 2011
Robert Koch Institute (2010) KiGGS - The German Health Interview and Examination Survey for Children and Adolescents ; Continuation of the KiGGS Study - wave 1 (2009–2012). http://www.kiggs.de/service/english/index.html. Accessed 28 Mar 2011
Robine JM, Ritchie K (1991) Healthy life expectancy: evaluation of global indicator of change in population health. BMJ 302:457–460
Rothgang H, Iwansky S, Müller R, Sauer S, Unger R (2010) BARMER GEK Pflegereport 2010. Asgard, St. Augustin
RTI International (2009) Methodology: “America’s Best Hospitals”. http://www.rti.org/pubs/abchmethod_2009.pdf. Accessed 30 July 2013
RTI International (2010) U.S. News & World Report Best Children’s Hospitals 2010 Methodology. http://www.rti.org/pubs/abchmethod_2010.pdf. Accessed 30 July 2013
Sauer K, Kemper C, Kaboth K, Glaeske G (2010) BARMER GEK Heil- und Hilfsmittel-Report 2010. Asgard, St. Augustin
Schäfer T, Nolde-Gallasch A (1999) Modellversuch zur Beitragsrückzahlung bei den AOKs Lindau und Ostholstein – Bericht der wissenschaftlichen Begleitung. Technical report, University of Applied Science Gelsenkirchen, Bocholt (forthcoming, Federal Association of the AOK, Bonn)
Schäfer T, Neusser S, Lorenz C, Dörning H, Bitzer EM (2007) Krankenhaus-Rangfolgen nach Ergebnisqualität in der Hüftendoprothetik – Routinedaten mit oder ohne ergänzende Patientenbefragungen? – Teil 1: Routinedaten. GMS Med Inform Biom Epidemiol 3(1):Doc08
Schäfer T, Schneider A, Mieth I (2011) BARMER GEK Zahnreport 2011. Asgard, St. Augustin
Schneeweiss S, Schöffski O, Selke GW (1998) What is Germany’s experience on reference based drug pricing and the etiology of adverse health outcomes or substitutions? Health Policy 44:253–260

Schneeweiss S, Seeger J, Maclure M, Wang P, Avorn, J, Glynn RJ (2001) Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. Am J Epidemiol 154:854–864
Schneeweiss S, Walker AM, Glynn RJ, Maclure M, Dormuth C, Soumerai SB (2002) Outcomes of reference pricing for angiotensin-converting-enzyme inhibitions. N Engl J Med 346:822–829
Schwabe U, Paffrath D (eds) (2010) Arzneiverordungsreport 2010. Springer, Berlin/Heidelberg
Schwartz FW, Busse R (2003) Denken in Zusammenhängen: Gesundheitssystemforschung. In: Schwartz FW, Badura B, Busse R, Leidl R, Raspe H, Siegrist J, Walter U (eds) Public Health. Gesundheit und Gesundheitswesen. Urban & Fischer, München/Jena, pp 518–545
Schwartz FW, Schach E (1989) Summary. In: Zentralinstitut für die kassenärztliche Versorgung in der Bundesrepublik Deutschland (ed): Die EvaS-Studie. Eine Erhebung über die ambulante medizinische Versorgung in der Bundesrepublik Deutschland. Deutscher Ärzte Verlag, Köln, pp 31–42
Schwarze E-W, Pawlitschko J (2003) Autopsie in Deutschland: Derzeitiger Stand, Gründe für den Rückgang der Obduktionszahlen und deren Folgen. Dtsch Arztebl 100:A2802–2808
Scott I, Campbell D (2002) Health services research: what is it and what does it offer? Intern Med J 32:91–99
Selby JV (1994) Case-control evaluations of treatment and program efficiency. Epidemiol Rev 16:90–101
Selmer RM, Kristiansen IS, Haglerød A, Graff-Iversen S, Larsen HK, Meyer HE, Bønaa KH, Thelle DS (2000) Cost and health consequences of reducing the population intake of salt. J Epidemiol Community Health 54:697–702
SF-36 Health Survey Scoring Demonstration. http://www.sf-36.org/demos/SF-36.html. Accessed 29 Mar 2011
SF-36 Psychometric Considerations. http://www.sf-36.org/tools/sf36.shtml. Accessed 29 Mar 2011
Shen Y, Ellis RP (2002) How profitable is risk selection? A comparison of four risk adjustment models. Health Econ 11:165–174
Shojania KG, Showstack J, Wachter RM (2001) Assessing hospital quality: a review for clinicians. Eff Clin Practice 4:82–90
Sonnenberg A, Delco F (2002) Cost-effectiveness of a single colonoscopy in screening for colorectal cancer. Arch Intern Med 162:163–168
Sörensen HT (2001) Routine registries. In: Olsen J, Saracci R, Trichopoulos D (eds) Teaching epidemiology. Oxford University Press, Oxford/New York, pp 99–106
Soumerai SB, Avorn J, Ross-Degnan D, Gortmaker S (1987) Payment restrictions for prescription drugs under Medicaid. N Engl J Med 317:550–556
Soumerai SB, Ross-Degnan D, Avorn J, McLaughlin JT, Choodnovskiy I (1991) Effects of Medicaid drug-payment limits on admission to hospitals and nursing homes. N Engl J Med 325:1072–1077
Spilker B (ed) (1995) Quality of life and pharmacoeconomic clinical trials. Lippincott-Raven, Philadelphia, PA
Statistisches Bundesamt (ed) (2000a) Gesundheitsbericht für Deutschland. Metzler-Poeschel, Stuttgart
Statistisches Bundesamt (ed) (2000b) Konzept einer Ausgaben- und Finanzierungsrechnung für die Gesundheitsberichterstattung des Bundes. Metzler-Poeschel, Stuttgart
Strauss DJ, Shavelle RM (2000) University of California life expectancy project. http://www.lifeexpectancy.com/index.shtml. Accessed 29 Mar 2011
Stuart A (1968) Basic ideas of scientific sampling. Charles Griffin & Co, London
Sudgen R, Williams A (1990) The principles of practical cost-benefit analysis. Oxford University Press, Oxford
Sukhatme PV, Sukhatme BV (1970) Sampling theory of surveys with applications. Asia Publishing House, London
Sullivan DF (1971) A single index of morbidity and mortality. HSMHA Health Rep 86:347–355
Sutton JH (2001) Physician data profiling proliferates. Bull Am Coll Surg 86:20–24

Tenney JB, White KL, Williamson JW (1974) National Ambulatory Medical Care Survey: background and methodology. In: National Center for Health Statistics (ed) Vital and health statistics series 2: data evaluation and methods research No. 61. U.S. Government Printing Office, Washington, DC

Thiemann DR, Coresh J, Oetgen WJ, Powe NR (1999) The association between hospital volume and survival after acute myocardial infarction in elderly patients. N Engl J Med 340:1640–1648

Torrance GW (1976) Social preference for health status. SocioEcon Plan Sci 10:129–136

Torrance GW (1987) Utility approach to measuring health-related quality of life. J Chronic Dis 40:593–600

Torrance GW, Thomas WH, Sackett DL (1972) A utility maximisation model for evaluation of health care programs. Health Serv Res 7:118–133

Townsend (2001) NHS resource allocation review: targeting poor health (vol I). In: Welsh Assembly's National Steering Group on the Allocation of NHS Resources. National Assembly for Wales, Cardiff

Tseng H-M, Lu J-F R, Gandek B (2003) Cultural issues in using the SF-36 Health Survey in Asia: results from Taiwan. Health and quality of life outcomes 1:72. http://www.hqlo.com/content/1/1/72. Accessed 29 Mar 2011

Tucker MA, Weiner JP, Abrams C (2002) Health-based risk adjustment: application to premium development and profiling. In: Wrightson C (ed) Financial strategy for managed care organizations: rate setting, risk adjustment, and competitive advantage. Health Administration Press, Chicago, pp 165–225

Tudor-Hart J (1971) The inverse care law. Lancet 1:405–412

Tudor-Hart J (2000) Commentary: three decades of the inverse care law. BMJ 320:18–19

U.S. Census Bureau (2009) Income, poverty and health insurance coverage in the United States: 2008. http://www.census.gov/newsroom/releases/archives/income_wealth/cb09-141.html. Accessed 29 Mar 2011

Van de Ven WPMM, Ellis RP (2000) Risk adjustment in competitive health plan markets. In: Culyer AJ, Newhouse JP (eds) Handbook of health economics, vol 1A. Elsevier/North Holland, New York, pp 755–845

van Mosseveld CJPM (2003) International comparison of health care expenditure. Statistics Netherlands, Voorburg/Herlen

Voß W (ed) (2003) Taschenbuch der Statistik, 2nd edn. Carl Hanser, München/Wien

Ware JE, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 30:473–483

Weinermann JE (1971) Research on comparative health services systems. Med Care 9:272–290

Weinstein MC, Stason WB (1977) Foundations of cost-effectiveness analysis for health and medical practices. New Engl J Med 296:716–721

Wennberg JE, Freeman JL, Culp WJ (1987) Are hospital services rationed in New Haven or over-utilised in Boston? Lancet 1:1185–1189

Wennberg JE, Freeman JL, Shelton RM, Bubolz TA (1989) Hospital use and mortality among Medicare beneficiaries in Boston and New Haven. N Engl J Med 321:1168–1173

White T, Lavoie S, Nettleman, MD (1999) Potential cost savings attributable to influenza vaccination of school-aged children. Pediatrics 103:73e

Williams A (1974) "Need" as a demand concept (with special reference to health). In: Culyer A (ed) Economic policies and social goals. Martin Robertson, London

Williams A (2001) Science or marketing at WHO? A commentary on 'World Health 2000'. Health Econ 10:93–100

World Health Organization (WHO) (1947) The constitution of the World Health Organization. WHO Chron 1:6–24

World Health Organization (WHO) (2000) World health report 2000. Health systems: Improving performance. WHO, Geneva

Wright J (2001) Assessing health needs. In: Pencheon D, Guest C, Melzer D, Muir Gray JA (eds) Oxford handbook of public health practice. Oxford University Press, Oxford, pp 38–46

Wüthrich-Schneider E (2000) Patientenzufriedenheit – wie verstehen? Teil 1. Schweizerische Ärztezeitung/Bulletin des médecins suisses/Bolletino dei medici svizzeri 81(20):1046–1048

Zhao Y, Ellis RP, Ash AS, Calabrese D, Ayanian JZ, Slaughter JP, Weyuker L, Bowen B (2001) Measuring population health risks using inpatient diagnoses and outpatient pharmacy data. Health Serv Res 26(6 Part II):180–193

Ziese T, Neuhauser H, Kohler M, Rieck A, Borch S (2003) Flexible Ergänzung der Gesundheitssurveillance in Deutschland: Gesundheitssurveys per Telefon. Gesundheitsberichterstattung des Bundes, Berlin. http://www.sgw.hs-magdeburg.de/kurmat/goepel/hoge/ggf/grundlagen/yhtml/pdf/tel-survey.pdf. Accessed 29 Mar 2011