

Describing and evaluating health systems

Ellen Nolte, Martin McKee and Suzanne Wait

Introduction: what is a health system?

Before one can even begin to discuss evaluation of health systems, it is first necessary to decide what a health system is. There is, unfortunately, no simple answer. A pragmatic view interprets a health system as being 'made up of users, payers, providers and regulators [that] can be defined by the relations between them' (McPake *et al.* 2002), with 'relations' referring to four key functions of health systems: regulation, financing, resource allocation and provision of services (Mills and Ranson 2001). In practice, however, health care systems are often defined by national borders, exemplified by the remark made frequently by journalists since the publication of the *World Health Report* (WHO 2000) that 'the French health care system [is] judged by the World Health Organization . . . to be the best in the world' (British Broadcasting Corporation 2000). Yet within each country there is almost always a complex mixture of different systems, in which some people use different ways to pay for health care and in turn receive different benefits (McKee and Figueras 1997). For example, while many people would identify the British health system with its National Health Service (NHS), a system established in 1948 to provide universal coverage paid from general taxation, that interpretation would miss the growing differences in the way in which health care is organized in the four constituent parts of the UK, with Scotland, in particular, moving increasingly away from the model evolving in England. Similarly, it would miss the substantial volume of health care provided in the private health care sector, both to those that have private health insurance and, increasingly, for those who choose to pay directly. And the UK is, in comparison with some countries, remarkably homogeneous. What, for example, is meant by the term 'American health care system', with its myriad of payment plans for those in employment, superimposed upon Medicare, for the elderly, and Medicaid (with its many variations from state to state) for the poor, to say nothing of a range of other federally funded programmes such as those for the armed forces, for veterans and for native Americans? Even the Soviet health care system, which might be thought to have been more homogenous than most, contained a large number of parallel systems for those employed in the armed forces, the railways, Aeroflot (the Soviet airline), as well as the *nomenklatura* (the Communist Party elite).

Then there is the problem of defining the boundaries of a health system. There are many activities that contribute, directly or indirectly, to the provision of health

care that, in different countries, may or may not be within what is considered to be the health system. The most obvious example is social care, especially for elderly people where it may be difficult, and indeed often inappropriate, to disentangle the provision of active health care (such as the investigation and treatment of chronic disorders) from more basic nursing care, to provision of appropriate living conditions. However, as health systems become increasingly complex, they depend ever more on a wide range of activities to generate and disseminate the knowledge required for effective health care, including basic and applied researchers and a growing body of 'knowledge brokers'. All contribute to the delivery of health care, yet they may be located in universities, industry, other branches of government, or one of the many charitable foundations working in the field of health. Similarly, does one include those involved in training health professionals? This role has often been linked closely with the provision of health care but, while remaining so, the nature of the association is changing. For example, in the UK, nurse training was until recently carried out by major hospitals but is now based in universities. The Soviet Union removed medical training from the universities in the countries of central and eastern Europe, placing them in institutions under the control of ministries of health, a policy that was reversed in many countries during the 1990s. Then there is the production, regulation and distribution of pharmaceuticals and medical technology, which like the training of professionals has, in some countries, moved across the interface that is commonly seen as the boundary of the health system, in particular in relation to products such as vaccines.

Yet it is not only diversity within the nation state that must be accommodated. Some countries operate health systems beyond their borders, most obviously in respect of troops deployed abroad but also, in a globalizing world, by corporations based in industrialized countries providing for their employees in other parts of the world. These may, *de facto*, owe more to the norms of the country from which they originate rather than the one in which they are located. Yet this is only one small effect of the process of globalization that – facilitated by agreements such as the General Agreement on Trade in Services that enable international corporations to move into the mainstream of health care delivery (Pollock and Price 2003) – means that the link between the nation state and the services it provides for its citizens becomes ever more tenuous.

Given this complexity, it is difficult to argue with Field's contention that the 'question of the drawing of the precise boundaries of [the health] system is an empirical and definitional one, and must, to some degree, remain arbitrary' (Field 1973). This, inevitably, leads to a situation in which different analysts choose different definitions. Thus, Anderson takes a narrow perspective, placing a health care system within the 'boundaries of a relatively easily defined system with entry and exit points, hierarchies of personnel, types of patients' (Anderson 1972). This health care system is 'the officially and professionally recognised "helping" services regarding disease, disability, and death'. More expansively, Field defines the health system as 'the aggregate of commitments and resources (human, cultural, political, and material) any society devotes to, or sets aside to, or invests into the "health" concern as distinguished from other concerns such as general education, defence, industrial production, communications, capital construction, and so on' (Field 1973). Yet he faces the problem of operationalizing this concept and, when developing it further using a structural-functional perspective, he proposes a more specific definition as 'that societal mechanism which transforms generalised resources or inputs (mandate, knowledge, personnel and resources) into specialised outputs in the form of health services aimed at the health problems of the society', with the 'health problems' being referred to as the five Ds: death, disease, disability, discomfort and dissatisfaction. A similar line of reasoning is followed by Roemer, who has arguably

written more on health systems than any other individual, and who defines the health system as 'the combination of resources, organization, financing, and management that culminate in the delivery of health services to the population' (Roemer 1991). Yet both these authors define the health system in terms of the structures used to deliver health care. In contrast, Weinerman, drawing on the World Health Organization (WHO) definition of health as the 'state of complete physical, mental and social well-being and not merely the absence of disease or infirmity' (WHO 1948), defined the health system as 'any set of arrangements in a society . . . which assigns social roles and resources to achieve the goals of protecting or restoring health to the eligible population' (Weinerman 1971). Although his analysis, in practice, focuses mainly on personal health services, this definition embraces 'all of the activities of a society which are designed to protect or restore health, whether directed to the individual, the community, or the environment'. In a similar vein, Long argues that, if health is to be interpreted in accordance with the WHO definition, then 'any service designed to improve the physical, mental, or social well-being of one individual or groups of individuals must be considered a health service' (Long 1994). Consequently, health care also includes education, housing, nutrition, environmental monitoring and others. However, Long also takes issue with the common practice of using the term *health care* interchangeably with *medical care*; instead he defines medical care as being only one of several types of services identified as health care services. Hence, the medical care system – as opposed to the health care system – refers to the organization, financing and delivery of medical care services that comprise three major generic components: preventive care, acute care and long-term care (Long 1994). In this respect, Long's definition is actually rather narrow, focusing on the 'health care system' solely as a provider of health services.

In 1998, WHO began to develop its Health System Performance Assessment Framework (HSPAF). This led to the publication of the *World Health Report 2000*, which was the first attempt to provide a comprehensive assessment of the performance of health systems in the then 191 member states of WHO (Murray and Frenk 2000; WHO 2000). This approach adopted a very broad definition of what constitutes a health system. It considered that the crucial determinant of whether something is within or outside a health system is the intent to improve health. It includes 'all actors, institutions and resources that undertake health actions – where the primary intent of a health action is to improve health . . . It incorporates selected intersectoral actions in which the stewards of the health system take responsibility to advocate for improvements in areas outside their direct control, such as legislation to reduce fatalities from traffic accidents' (Murray and Evans, 2003). With this, WHO has arrived at one of the major challenges facing those seeking to evaluate health systems: even if one can reach a satisfactory definition of what a health system actually is, how does one disentangle its effects from the many other things that are taking place within the society in which it is embedded?

Yet there is another problem to be addressed. A frequent reason for assessing the performance of a health system or sub-system is to draw lessons from that assessment. Yet health systems exhibit strong path dependency. Many of the national specificities of each health system are determined by particular historical circumstances, such as the emergence of western European social insurance systems from strong sets of relationships between employers and employee associations in Germany and France following the industrial revolution, the rejection of centralized state control in the countries of central and eastern Europe that emerged from communist rule in the 1990s, the shared wartime experience that led to the creation of the British NHS, or the rugged individualism and non-conformism that characterizes much of American life, and by extension the delivery of health care. As a

How

consequence, most analysts recognize that health systems cannot simply be relocated from one country to another (although unfortunately this understanding does not always extend to politicians and their advisers).

In summary, different authors have, at different times, employed quite different definitions of what a health system is. The lesson that can be drawn is that, whatever definition is being used, it is essential that it be defined explicitly and the means of evaluating this system are congruent with the definition. Yet beyond the question of which national system is best, there is the question of whether one type of system, such as one funded from taxation (often characterized as a 'Beveridge', system after the British architect of that country's NHS) or one funded by social insurance (often characterized as 'Bismarckian', after the German chancellor who introduced it in the latter part of the eighteenth century) is superior. To address this question it is first necessary to understand the various ways that have been used to classify health systems.

How does one classify health care systems?

For years, health policy researchers have asked 'Can one develop a classification of health care systems?' The way in which this quest has been pursued provides valuable insights into the difficulties involved and, in particular, the dangers of simplification.

Many of the most simple classifications, such as that containing the Bismarck and Beveridge models mentioned above, are derived from the concept used by Max Weber of 'ideal types' (Weber 1950). An ideal type refers to *an abstract model of a complex real phenomenon, which highlights its most significant features*. In this context, 'ideal' is not meant in the sense of desirable but in the sense of a pure, abstract construct, going back to the Platonic view that what one sees on earth is an imperfect representation of something that exists in some ideal world. This approach offers a series of hypothetical models that emphasize certain features that may have some explanatory power. Such models often reflect some underlying view about the way in which society is organized. It should also be noted that much of the literature that has adopted this perspective is concerned, at least implicitly, with one question, which has thus shaped its application; why, among industrialized countries, is the USA unique in not having developed a system of universal health care coverage?

One example is that developed by Field, who identified five ideal-type health systems that reflect the diversity of different patterns of health care organization (see Table 2.1) (Field 1978). In this typology, the key dimensions that define a health care system include the role of the state versus that of the market, as well as the position of the physician, the role of professional associations and the ownership of facilities.

An analogous approach is that developed by Roemer, who proposed a typology of health systems on two dimensions: the level of economic development, classified according to the gross national product (GNP), and political characteristics, namely the level of market intervention in health policy (see Table 2.2) (Roemer 1977, 1991). In this two-dimensional matrix, each dimension consists of four (originally three – Roemer 1977) levels, with the economic dimension distinguishing between 'affluent and industrialized', 'developing and transitional', 'very poor' and 'resource-rich'. The political categories include 'entrepreneurial and permissive', 'welfare-oriented', 'universal and comprehensive' and 'socialist and centrally planned'. Illustrative examples include an entrepreneurial system in an industrialized country,

Table 2.1 Types of national health systems, as classified by Field

	<i>General definition</i>	<i>Position of physician</i>	<i>Role of professional associations</i>	<i>Ownership of facilities</i>	<i>Economic transfers</i>	<i>Prototypes</i>
<i>Type 1</i> Private	Health care as item of personal consumption	Solo entrepreneur	Powerful	Private	Direct	USA, Western Europe
<i>Type 2</i> Pluralistic	Health care as consumer good or service	and member of variety of groups/organizations	Very strong	Private and public	Direct and indirect	USA in twentieth century
<i>Type 3</i> National health insurance	Health care as an insured/guaranteed consumer good or service	and member of medical organizations	Strong	Private and public	Mostly indirect	Sweden, France, Canada
<i>Type 4</i> National health service	Health care as a state-supported consumer good or service	and member of medical organizations	Fairly strong	Mostly public	Indirect	Great Britain
<i>Type 5</i> Socialized health service	Health care as a state-provided public service	State employee and member of medical organizations	Weak or non-existent	Entirely public	Entirely indirect	Soviet Union

Source: adapted from Rodwin (1984)

Table 2.2 Types of national health systems, as classified by Roemer

		<i>Health system policies (market intervention)</i>			
		<i>Entrepreneurial & permissive</i>	<i>Welfare-oriented</i>	<i>Universal & comprehensive</i>	<i>Socialist & centrally planned</i>
<i>Economic level (GNP/capita)</i>	<i>Affluent & industrialized</i>	USA	West Germany Canada Japan	Great Britain New Zealand Norway	Soviet Union Czechoslovakia
	<i>Developing & transitional</i>	Thailand Philippines South Africa	Brazil Egypt Malaysia	Israel Nicaragua	Cuba North Korea
	<i>Very poor</i>	Ghana Bangladesh Nepal	India Burma	Sri Lanka Tanzania	China Vietnam
	<i>Resource-rich</i>		Libya Gabon	Kuwait Saudi Arabia	

Source: "Figure 4.1", from NATIONAL HEALTH SYSTEMS OF THE WORLD, VOLUME I: THE COUNTRIES by Milton I. Roemer, copyright © 1991 by Oxford University Press, Inc. Used by permission of Oxford University Press, Inc.

such as the USA, a welfare-oriented system in a transitional country such as Brazil, and a socialist system in a poor country, as exemplified by Vietnam.

Arguing from a political-economic perspective and drawing on a Marxist interpretation, Elling (1994) proposed classifying countries' health systems in order of increasing strength of their labour movements. This yields five types of countries, and thus health systems:

- 1 core capitalist;
- 2 core capitalist, social welfare;
- 3 industrialized socialist-oriented;
- 4 capitalist dependencies;
- 5 socialist-oriented, quasi-independent.

Thus, *core capitalist* countries are characterized by low strength of workers' movements, a decentralized, fractionated authority structure, a market-oriented health system that may include elements of a national insurance system, and gross disparities in distribution of wealth, access to health services and levels of health in terms of class, ethnicity or gender. Examples include the USA, Switzerland and Germany. The second type, *core capitalist – social welfare*, includes countries with stronger workers' movements and a better developed welfare system, with either a regional or national health (insurance) system. Examples include Canada, the UK and the Scandinavian countries. The third type, *industrialized socialist-oriented*, has largely disappeared with the break up of the Soviet Union, with the most prominent features being that the workers' movements were subsumed within the Communist Party, there were fewer social and economic disparities than in types (1) and (4) and there were partially (administratively) regionalized national health services. *Capitalist dependencies* are characterized by the workers' movements being suppressed, with little or no collective provision of health and welfare services and 'obscene social and economic as well as healthy disparities' (Elling 1994) as in Brazil, India and the Philippines. Finally, the main features of the *socialist-oriented – quasi-independent* type include strong workers' and peasants' movements, regionalized health services and greater equity in the distribution and control of resources including health services (e.g. China, Cuba, Tanzania).

These approaches are purely illustrative as other writers have developed their own typologies, although most are variations on the same themes (e.g. Maxwell 1974; Terris 1978; Raffel 1984). From a contemporary perspective, as Sheaff notes, they largely reflect 'certain political preoccupations of [the cold war] time. Then, a touchstone of political and social analysis was where a society or an economic sector fell in terms of the global political division between fundamentally market-based and fundamentally state-managed social systems' (Sheaff 1998).

While political scientists continue to debate whether the world is unipolar (i.e. dominated by the USA) or multipolar, what is incontrovertible is that the world is no longer divided into two competing camps, capitalism and communism. As a consequence, 'taxonomies reflecting Cold War alignments have become unrealistically narrow' (Sheaff 1998) and ignore the multi-dimensionality that characterizes the provision of health care, a point developed by more recent commentators.

An example is the model developed by Frenk and Donabedian (1987). Rather than providing a typology of health systems, they developed a typology based on certain configurations of state intervention in health care in relation to specific principles for the population's eligibility to receive care. The original model focused on the supply side of services, which was categorized according to, first, the degree of ownership – whether the state limits its role to the financing of care or also assumes the role of a health care provider – and, second, the administrative structures, reflecting the concentration of control – i.e., is control concentrated in a

Table 2.3 Typology of health care modalities

		<i>Basis for population eligibility</i>			
		<i>Purchasing power</i>	<i>Poverty</i>	<i>Socially perceived priority</i>	<i>Citizenship</i>
<i>Mechanism for state intervention (degree of control)</i>	<i>Regulation</i>	Private enterprise	Private charity	Company-based services	Social insurance (German model)
	<i>Financing</i>	–	Medicaid (USA)	Incipient health insurance	National health insurance
	<i>Delivery</i>	–	Public assistance	Social security (Latin American model)	Socialized (national health service)

Reprinted from Health Policy, v.27:19–34, Frenk: Dimensions of health system reform, c 1994, with permission from Elsevier.

single agency or programme or is it dispersed among several agencies. This model was subsequently expanded to also include aspects of financing and regulation (see Table 2.3) (Frenk 1994).

For example, a characteristic feature of the German model is that financing is operated by private, non-profit funds that contract private providers. The role of the state is largely restricted to regulating these groups and to establishing a regulatory framework that guarantees minimum levels of benefits to which all citizens are entitled. In contrast, in countries such as the UK and Sweden, the state has been responsible for the delivery of most services.

This approach also makes it possible to disaggregate the various modalities of state intervention that may coexist in any given country, such as the multiple elements of the American system. Thus, company-based services, under which private employers organize the financing and delivery of services for their workers, exist alongside state financing of health services for the poor (Medicaid) alongside state provision of services to particular sub-groups of the population (e.g. veterans).

Another approach is to step down a level further to classify countries on the basis of more specific aspects of their health system. Thus, in 1992, the Organization for Economic Cooperation and Development (OECD) undertook a systematic analysis of health care systems that sought to identify the dominant mechanisms for funding, payment and regulation in seven OECD countries in western Europe (OECD 1992). It drew on earlier work by Evans (1981) who proposed distinct models that summarized interactions between five principal sets of actors in health care systems: (a) consumer/patient, (b) first-level providers (e.g. general practitioners, pharmacists supplying over-the-counter medicines), (c) second-level providers (e.g. hospital services, pharmacists supplying prescribed drugs), (d) insurers (or third-party payers) and (e) government in its capacity as regulator of the system. The main interactions include provision of services, referrals from first- to second-level providers, payment for services, payment for insurance, payment of insurance claims and various forms of regulation by government. Using this model, the authors then identified seven models to describe the sub-systems of finance and methods of paying providers (see Table 2.4). These models are further illustrated with diagrams depicting financial and patient flows and the relationships between patients, providers and third parties, in each case following a standardized, highly structured format.

In this, and in its other work, in particular in developing national health accounts,

Table 2.4 Sub-systems of finance and provider payments

<i>Model</i>	<i>Example</i>
Voluntary, out-of-pocket payment	Supporting role only, e.g. purchase of over-the-counter medicines, cost-sharing for prescribed medicines
Voluntary (insurance with) reimbursement of patients	Private sector in UK and Netherlands
Public (compulsory insurance with) reimbursement of patients	Elements retained in the social health insurance systems in France and Belgium
Voluntary (insurer/provider) contract	Individual Practice Association and prepaid group practices in Spain (private sector)
Public (insurer/provider) contract	Primary care in Germany, Netherlands, Ireland, UK; hospitals in Belgium, Germany, Netherlands, UK
Voluntary insurance with integration between insurers and providers	USA: Health Maintenance Organizations
Compulsory insurance with integration between insurance and providers	Spain; public hospitals in France and Ireland (previously public hospitals in UK)

Source: OECD (1992)

the OECD has made important contributions to the comparison of health systems, not least in highlighting the need for a systematic approach and, in particular, the use of agreed definitions. However it also illustrates the complexity involved, as this classification based on systems of financing deals solely with revenue, while a classification based on capital financing (e.g. funding and ownership of hospitals and other health care facilities) would look quite different (Thompson and McKee 2004).

A completely different approach has emerged from work on complex adaptive human systems, based on soft systems theory (Checkland 1981). This approach implicitly rejects the concept of ideal types and sees the health system, like any other system, as somewhat more complicated. The health system is a complex 'whole' that is made up of a hierarchy of levels of organization, or sub-systems, with higher levels becoming progressively more complex. According to Checkland (1981), the leading exponent of soft systems theory, a system has certain features:

- it has a purpose or mission and its performance can be measured;
- it contains decision-making processes that are themselves systems and these interact so that their effects can be transmitted throughout the system;
- it exists in wider systems and/or environments with which it interacts but from which it is separated;
- it has resources that can be used by the decision-making process;
- it has some degree of continuity.

Furthermore, unlike the implication of some other approaches that also break health systems into their constituent parts, this approach rejects the idea that the characteristics of a system, analysed on a given level, can be predicted from knowledge of the sub-systems that contribute to it, as each level displays emergent properties that do not exist at lower levels. An analogy is that of a living organism, such as a human in which, as identical twins demonstrate, even a complete knowledge of the constituent genes does not allow the investigator to predict with certainty all the characteristics of the individual twin.

For the analyst, the key issue is that the appropriate level at which evaluation should take place is determined by the question being asked. Each level of complexity is characterized by specific features that require specific approaches and techniques for analysis. Wilson and Holt (2001) illustrated this, although not explicitly referring to soft systems theory, in relation to human beings who, they argue, can be considered as composed of and operating within multiple interacting and self-adjusting systems. Looking at human health and illness, they identify several levels of 'systems', each requiring specific approaches of analysis. The human body, for example, is composed of multiple interacting and self-regulating physiological systems whose interactions and functioning can be investigated by using a variety of biochemical and physiological techniques. The next level is the behaviour of the individual that is determined by a complex set of rules based on past experience and responses to environmental stimuli whose complexity may be understood more closely by applying techniques derived from psychology and related disciplines. The set of rules and experiences determining individual behaviour itself is largely influenced by relationships the individual is embedded in and which impact their beliefs and expectations. Appropriate methods to understand these interdependencies would be derived primarily from the social sciences, such as social psychology. However, individuals and their immediate social relationships are further embedded within wider social, political and cultural systems that 'can influence outcomes in entirely novel and unpredictable ways' (Wilson and Holt 2001). Potential approaches to interpreting this level of complexity would involve a variety of disciplines including anthropology, social sciences, political sciences and economics.

It should, however, be recognized that, rather like complexity theory, which has been shown to explain such diverse phenomena as the pattern of migrating birds, the population of wild animals and the behaviour of stock markets (Lewin 1992), soft systems analysis suffers from a major limitation, and one that diminishes it in the view of many politicians: it cannot *predict* what will happen. A health system, like a living organism, contains processes of communication and control that enable it to adapt in response to environmental pressures. In other words, it cannot be assumed that an intervention that was successful in one setting will necessarily work in another. Whether such outcomes are actually predictable is, of course, another matter (McKee 1995).

As both the OECD model and the applications of soft systems theory show, there has been a move away from the evaluation of the system as a whole (with the notable exception of the 2000 *World Health Report*) to assessments of different ways of achieving some of the many functions that contribute to the overall health system or, put another way, to the evaluation of sub-systems.

Getting inside the system: a framework for assessment

The levels within a health care system are potentially almost infinite, reflecting the very many questions that it is possible to ask about a system and its components, and taking account of the many problems in defining the boundaries of the system discussed earlier. One simplified approach is to look at the different levels of decision-making within a health care system: the *primary process of patient care* (micro level); the *organizational context* (meso level); and the *financing and policy context* (macro level) (Ploch and Klazinga 2002). Each level is characterized by distinct rationales, addressing different dynamics in the health care system; for each level it is thus possible to identify specific issues that ultimately shape the health care system, for example:

- *Micro level* – what is the nature of the interaction between health service users and professionals?
- *Meso level* – what is the most effective balance between inpatient and ambulatory (outpatient) care?
- *Macro level* – how are health services financed?

The model underpinning the HSPAF, as set out in the *World Health Report 2000*, is more complex (WHO 2000). As already noted, the framework identifies three major social goals to which health systems contribute, namely health attainment, responsiveness to the expectations of the population and fairness of financial contribution. However, in order to achieve these goals or objectives the health system has to fulfil certain key functions; these are identified as *financing*, *provision* of personal and non-personal health services, *resource generation* and *stewardship*, or the oversight function of the health system (see Figure 2.1).

Each function can be further divided into distinct sub-functions that can be analysed separately. Thus, financing involves the components revenue collection, fund pooling and purchasing (see Box 2.1). In brief, revenue collection refers to the process of mobilizing resources (i.e. money), usually from households or corporate entities but also from governments and external donors. Fund pooling refers to the spreading of financial risk across the population through the accumulation of pre-paid health care revenues, while purchasing is the process through which revenues that have been collected are allocated to providers who must deliver a package of services.

Similarly, the function 'provision' can be subdivided into personal health services, i.e. services that are consumed directly by an individual, and non-personal health services, i.e. actions that are applied either to collectives (e.g. mass health education) or to the non-human components of the environment, such as basic sanitation. Box 2.1 illustrates the wider implications of the health system functions as outlined by the *World Health Report 2000*. Thus, it is important to stress that resource generation is much more than collecting money. It also involves forward planning to ensure there is something to buy with the money and which not only relates to human and physical resources but also to intellectual and social resources.

This approach offers a basis for categorizing the various elements within a health system. A next step is to describe how they operate. Here it is possible to derive

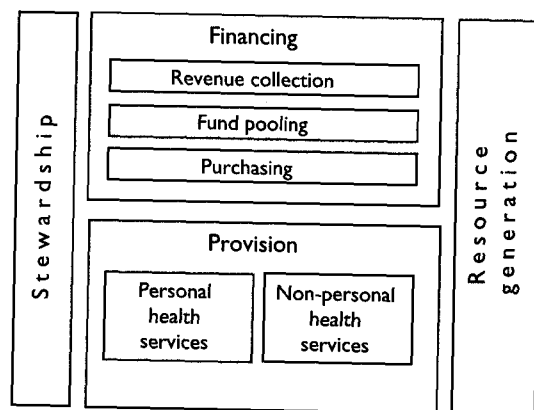


Figure 2.1 Functions of health systems

Source: Murray and Frenk (2000). Reproduced with kind permission of the publisher.

Box 2.1 Functions and sub-functions of health systems

Financing	<p><i>Revenue collection:</i> establishing prepayment systems and protecting the poor from catastrophic illness</p> <p><i>Fund pooling:</i> maintaining equity between generations, social groups etc.</p> <p><i>Purchasing:</i> making sure the money is used appropriately</p>
Provision	<p>Provision of care for the right people, in the right place, in the right way</p> <p>Balancing strategic purchasing with provider autonomy</p> <p>Linking quality with control over resources</p>
Resource generation	<p><i>Physical capital:</i> investment in high-quality, appropriate facilities</p> <p><i>Human capital:</i> investment in trained, motivated people with the appropriate mix of skills</p> <p><i>Intellectual capital:</i> investment in research and development</p> <p><i>Social capital:</i> investment in networks and relationships</p>
Stewardship	<p><i>Health policy formulation:</i> defining the vision and direction for the health system</p> <p><i>Regulation:</i> setting fair rules of the game with a level playing field</p> <p><i>Intelligence:</i> assessing performance and sharing information</p>

Source: adapted from Murray and Frenk (2000), World Health Organization (2000)

some insights from soft systems theory (Checkland 1981). This is based on an acronym – CATWOE – which describes any form of human activity and the circumstances that surround it (see Box 2.2). A certain *transformation* (process) is performed for clients (those who more or less directly benefit – *customers* – or suffer) by *actors*. The activity is ultimately ‘controlled’ or paid for by *owners*, and its implementation is influenced by the *environment* within which it is located. This all takes place against a background of various beliefs or values, in this case termed *Weltanschauung*, or world view.

For example, one might describe the British NHS as ‘a system for meeting the health needs of the entire population (*transformation, customers*) through the activities of those working in the NHS (*actors, implied ownership* by government), within limited resources (*environmental constraints*) and in the belief that health care free at the point of delivery is a good thing and most health professionals are essentially altruistic (*Weltanschauung*)’. Going down a level, its system of financing could be described as ‘a system for distributing money collected for health care to hospitals and health care workers (*transformation, customers*), in a way determined by government, advised by review bodies (*ownership, actors*), in the light of competing claims on government expenditure (*environmental constraints*), in the belief that rewards to

Box 2.2 CATWOE model

Customers	beneficiaries of the system
Actors	who carry out, or cause to be carried out, the transformation
Transformation process	the means by which defined inputs are transformed into defined outputs
Weltanschauung	vision of the world assumed for the system to function
Ownership of the system	someone with prime concern for it and the power to cause it to cease to exist
Environmental constraints	in the environment (geography, national wealth) or related systems (educational, legal, governmental, financial)

Source: Checkland (1981)

staff in the health care sector should be commensurate with those in other sectors (*Weltanschauung*).

While this provides a structured, systematic approach to describing health systems and the various elements that make them up, it does not say anything about how they are performing. At all levels of a health system, this can be assessed in terms of what is actually achieved, how it is achieved, and whether there are the prerequisites available for the system to achieve. Put another way, and adopting the approach first developed by Avedis Donabedian in his work on quality of care (Donabedian 1966, 1980), a system can be evaluated according to *structure*, *process* and *outcome*. Donabedian argued that 'good structure increases the likelihood of good process, and good process increases the likelihood of good outcome' (1988). The approach has subsequently been adopted widely within health services research and, in the specific context of evaluation of health systems, to include *outputs*, referring to the throughput or productivity of the health care system, i.e. the immediate result of professional or institutional health care activities, usually expressed as units of service (see Box 2.3) (Last 2001).

In summary, although it is common for media commentators (and some politicians) to speak of the British or the American or French health system, a rather more sophisticated approach is required. Several steps are needed. The first is to decide the precise nature of the question being asked. Is the subject of interest the overall health system, and if so, how are the boundaries of the system defined? If it is one element within the system, what is its purpose and what elements does it comprise? The framework set out in the *World Health Report 2000*, while not exhaustive, provides a useful starting point to think about the various elements that make up a health system. A second step is to describe the systems, or sub-systems being considered. The soft systems approach, using the CATWOE framework, may be of help here, not least because, in the area of comparative research, it will often highlight how like is not being compared with like. For example, during the 1990s, there was considerable enthusiasm from some politicians in the UK for the system of social insurance funding that exists in, for example, Germany. However, a simple application of this framework would have highlighted the very important role in Germany of employers' associations and trade unions, working through

Box 2.3 Dimensions of health services and health systems

• Structure (input)	<p>Attributes of the settings in which care occurs: resources needed for health care</p> <p>Material resources (facilities, capital, equipment, drugs etc.)</p> <p>Intellectual resources (medical knowledge, information systems)</p> <p>Human resources (health care professionals)</p>
• Process	<p>Use of resources: what is done in giving and receiving care</p> <p>Patient-Related (intervention rates, referral rates etc.)</p> <p>Organizational (supply with drugs, management of waiting lists, payment of health care staff, collection of funds etc.)</p>
• Output	<p>Productivity or throughput</p> <p>Length of stay in hospital, waiting times, discharge rates, access, effectiveness, equity of care</p>
• Outcome	<p>Effects of health care on the health status of patients and populations</p> <p>Definite: mortality, morbidity, disability, quality of life</p> <p>Intermediate: blood pressure, body weight, personal well-being, functional ability, coping ability, improved knowledge etc.</p>

well-established systems of industrial governance, a model that simply does not exist in the UK (Green *et al.* 2002).

Having defined the system of interest, the final step is to decide how to evaluate it. The model developed by Donabedian provides a basis for consideration, separating structures, processes and outcomes. The experiences of those undertaking evaluations of the performance of health systems will be examined later but, for now, it may be helpful to step sideways to review the history of international comparisons of health systems.

International comparisons of health systems

Learning about other countries is rather like breathing: only the brain dead are likely to avoid the experience.

(Klein 1997)

On any matter not self-evident, there are ninety-nine persons totally incapable of judging of it, for one who is capable.

(John Stuart Mill, *On Liberty*)

Interest in cross-national comparisons of health care systems can be traced back to the 1930s, with roots in an interest in the historical evolution of health care systems, as exemplified by the work of Sigerist (1943), much of which had the goal of informing developments in national health policy (Goldman 1946; Mountin and Perrott 1947). Cross-national comparisons received increasing attention from the 1960s onwards, the most influential examples being works by Abel-Smith (1963,

1967), Roemer (1960, 1969), Anderson (1963) and Mechanic (1975) to name but a few. Comprehensive overviews of work undertaken up to the 1960s and 1970s have been assembled by Weinerman (1971) and Elling (1980).

A key message of this chapter is that the approach taken in describing and analysing health systems depends critically on the question being asked. In judging what has been done previously, therefore, it is necessary to examine the background against which it took place. Much, though not all, of this research has its origins in the USA. This was a time when the economy was booming, and with it the health care system, in what Relman described as the 'era of expansion' (1988). Technological developments seemed to offer boundless possibilities, echoed in another area by the successful quest to place a man on the moon. However, successive extension of coverage of population groups in insurance-based systems or, as in the USA, the introduction of Medicare and Medicaid in the mid-1960s, giving more citizens access to care, also led to an increase in demand and consequently rapid growth in health expenditure in many industrialized countries, by then entering a period characterized by Relman as the 'era of cost containment': 'Increasingly, health administrators have been called upon to explain their demands for more and more national resources' (Abel-Smith 1967).

In part reflecting the availability of data but also the political concern about health care spending, much work that has been undertaken subsequently was mainly from a health economics perspective, looking mostly at health care expenditure and its determinants (Kanavos and Mossialos 1990). The most prominent examples include the work by the OECD since the 1980s in an effort to provide an empirical basis for a comparative understanding of the differences and similarities between OECD countries' health systems (OECD 1985; Schieber 1987). This emphasis on inputs into health care has changed only recently in the light of increasing pressures for reform of health care delivery, with many countries facing similar problems of rising costs, demographic changes, technological advances and increasing consumer expectations. There has been increasing interest in the possibility of learning from the many experiences of others, drawing lessons on how to finance, manage and organize health care so as to improve the overall performance of health systems. This last point has gained particular momentum on national and international agendas with the publication of *The World Health Report 2000* and its ranking of the world's health systems (WHO 2000), stimulating a wide-ranging debate about approaches to assessing health system performance both nationally and internationally (OECD 2002), which will be examined in more detail in the final section of this chapter.

Approaches to health system comparisons fall broadly into one of three main groups: descriptive studies, quantitative approaches and focused analytical studies.

Descriptive studies

Descriptive studies are systematic, structured descriptions of health systems or their sub-systems that can provide a basis for subsequent analysis. The use of a clear structure identifies areas that are unclear or poorly thought out. Examples include the work by the OECD described above (OECD 1992). The OECD reports provided a systematic assessment of the sub-systems of finance and methods of paying providers through the application of a standard and highly structured format.

This approach has been adopted by the European Observatory on Health Systems and Policies in its Health Care Systems in Transition (HiTs) documents (European Observatory on Health Systems and Policies, 2004), which provides a highly structured description of health care systems in Europe and other industrialized countries. Beginning with contextual information about the country, HiTs

describe the entities involved in financing, paying for and delivering care, drawing out their often complex interrelationships. HiTs then conclude with an examination of trends in health system reform. Prepared by a team that includes authors from the country in question as well as the Observatory, HiTs go beyond the formal structures to reflect the often messy reality of relationships. HiTs are now available for over 40 European countries, as well as some exemplar countries in other parts of the world, such as Australia and New Zealand. While not intended as a means of comparing systems, HiTs do contain a number of comparative tables, looking at each country's position in terms of, for example, resources used and outputs achieved (while noting the limitations of the data).

Another example is the International Network for Health Policy and Reform, which draws on information gathered from currently 16 industrialized countries, building on the presence of a partner institution in each country, and using a biannual survey of health reforms and health policy developments. The survey follows a highly structured format with standardized definitions and the information is drawn together in the form of regular published and online reports (International Network for Health Policy and Reform 2004).

Other approaches make use of the wealth of quantitative data collected in a fairly standardized format by international organizations such as the OECD (2003) and the WHO Regional Office for Europe (WHO Regional Office for Europe 2004). One example is the Commonwealth Fund programme on multinational comparisons of health systems data, which compares the US health care system with those in 28 industrialized countries in terms of, variously, financing, expenditures, availability and use of services, responsiveness to patients, and health outcomes. These are published on an annual basis (Reinhardt *et al.* 2002). Although relatively easy to do, such comparisons face the obvious problem of comparability. Some of the difficulties will be discussed in detail later, but, fundamentally, these approaches suffer from the problem that what can be counted is not necessarily what is important.

Quantitative approaches

Quantitative approaches have most often evolved from the health economics perspective to assess the performance of health systems in international comparison. There is a large literature on international comparisons of health expenditure, exploring the relationship between national wealth (such as gross domestic product, GDP) and health expenditure (Parkin *et al.* 1987; Kanavos and Mossialos 1990; Milne and Molana 1991). These studies do, however, yield conflicting results and it has been argued that, because of the considerable challenges involved in measuring health expenditure and national wealth, the observed positive relationship between health spending and GDP is unhelpful and likely to be misleading for health policy development (see Box 2.4).

Other studies have employed a production function approach that describes 'the production of health in terms of a function of possible explanatory variables' (Buck *et al.* 1999), usually examining factors indicative of health care ('health care input') and other explanatory variables for their impact on some health measure ('health care output') through regression analysis. Examples include a series of studies by the OECD that examined the associations of a number of input and process indicators such as health care expenditure, number of physicians, type of provider payment or access to services with health outcomes such as premature mortality and infant mortality (Or 2000, 2001). Other studies examined the association between specific aspects of health care systems and selected health outcomes – for example, the strength of the primary care system in different countries as a predictor for health outcomes (see Box 2.5) (Macinko *et al.* 2003).

Box 2.4 International comparisons of health expenditure: how valid are they?

Cross-country comparisons of health expenditure require adjusting expenditure according to the relative cost of what is being purchased. This is done by means of purchasing power parity (PPP) adjustment, in which the price of a basket of goods in each country is compared. First, it is important to specify whether general or health-specific PPPs are being used, as changes in the two are only imperfectly correlated. Furthermore, there are different PPPs to choose from, calculated by different organizations, such as the OECD and EUROSTAT, the statistical office of the European Union, each covering their own member states. Second, even when using health PPPs, it is important to recognize their limitations as they are only recalculated every five years and they focus largely on internationally traded goods, and in particular on pharmaceuticals, largely ignoring the major cost of staff in most health care systems.

Box 2.5 The contribution of primary care systems to health outcomes

Starfield and colleagues undertook a series of studies assessing the contribution of primary care systems to health outcomes in various settings. Defining primary care as 'that level of a health service system that provides entry into the system . . . provides person-focused care over time, provides care for all but very uncommon or unusual conditions, and coordinates or integrates care provided elsewhere or by others', one study looked specifically at the relationship between primary care and health outcomes in 18 OECD countries for the period 1970–98 (Macinko *et al.* 2003). The strength of primary care (PC) was measured using a ten-component scale reflecting structural characteristics, for example financing, resource allocation and accessibility, and specific practice features of PC, such as gatekeeper function, comprehensiveness and coordination. These components were then scored according to predefined criteria and combined to form a summary score, ranging from 0 (no component defined as characteristic for PC present) to 20 (all components present). In applying this model to 1995 data, France was shown to score lowest (2) and the UK highest (19). The relationship between PC strength and mortality as health outcome was then assessed using a regression model.

This showed the strength of a country's primary care system to be significant and negatively associated with all-cause (premature) mortality and premature mortality from selected conditions including asthma and bronchitis, emphysema and pneumonia and cardiovascular disease even after adjustment for a number of health determinants such as national wealth (GDP) or alcohol and tobacco consumption. Keeping some limitations of the analysis in mind, such as its ecological design and the limitations inherent in the underlying data, the overall findings suggest that the financing, organization and delivery of primary care seem to have an important impact on population health.

In addition to the limitations of the data, some of these studies are problematic because the theoretical basis of the relationships is not set out clearly, often giving the impression that the model was driven by data availability rather than plausible mechanisms and, especially when studies use a measure of adult mortality (such as life expectancy) as a dependent variable, they fail to take account of the well-known lag effects between exposures and outcomes that affect many disease processes.

The work on the performance of health systems by WHO, set out in the *World Health Report in 2000*, offers a somewhat different approach. Drawing on the goals of the health system as set out in its performance assessment framework, WHO used three main indicators to measure performance: population health, responsiveness and fair financing. Overall health system performance was then assessed as a composite of these indicators, which was in turn compared with what might be expected given the country's level of economic and educational development. The 191 WHO member states were then ranked according to these performance measures, producing a highly controversial league table of the world's health systems. The report played an important role in stimulating a wide-ranging debate on health system performance, and the various criticisms that it engendered helped bring to light the methodological challenges inherent in conducting and interpreting international comparisons, which are discussed below.

Yet another approach has evolved from epidemiology, involving analysis of data on mortality at a population level that are routinely available in many countries. It is based on the concept that certain deaths should not occur in the presence of timely and effective medical care (Nolte and McKee 2004). This concept of 'avoidable mortality' was introduced in the 1970s as a means to assess the quality of health care (Rutstein *et al.* 1976) and was subsequently adopted by a wide range of researchers especially in Europe (Charlton *et al.* 1983), producing for example the *European Community Atlas of 'Avoidable Death'* (Holland 1988). Much of this work dates back to the 1980s and early 1990s; only recently has this concept been revitalized as a potential useful tool to assess the quality and performance of health systems (see Box 2.6) (Nolte and McKee 2003, 2004).

Box 2.6 Avoidable mortality

Nolte and McKee (2004) have undertaken a systematic review of empirical and methodological studies of the concept of 'avoidable' mortality and of studies of the attribution of outcomes to health care. Their study demonstrated that 'avoidable' mortality proved a valuable instrument for the detection of potential weaknesses in health care that could then be investigated in more depth, and that it continues to do so.

Building on this work, the authors used a modified version of the concept of avoidable mortality by updating the list of conditions considered amenable to health care in the light of advances in medical knowledge and technology and looking at deaths under 75 years of age. The concept was then applied to routinely available data from selected countries in the European Union to investigate the potential impact of health care on changing life expectancy and mortality in the 1980s and 1990s. This showed that, since 1980, all European countries experienced increases in life expectancy between birth and age 75, although the pace of change differed over time and between countries. Reductions in amenable mortality made substantial

positive contributions in the 1980s in all countries. The largest contribution was from falling infant mortality but there were also improvements among the middle-aged, for example in Denmark, The Netherlands, the UK, France and Sweden. In many countries the pace of improvement slowed in the 1990s although not in the Mediterranean countries, a finding that would imply a continued catching up in the southern European countries. As a result, in the 1990s, differences in amenable mortality in the European Union had narrowed, although standardized death rates from amenable causes among Portuguese men remained three times higher than those among Swedish men (see Figure 2.2).

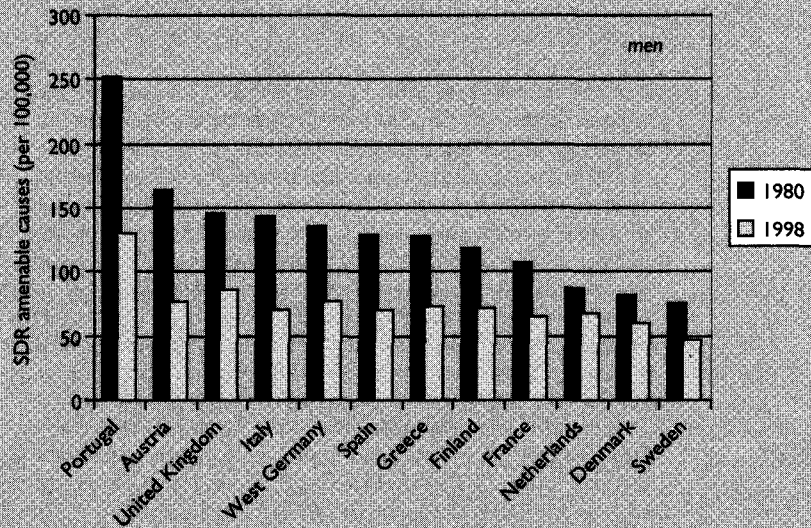


Figure 2.2 Age-standardized death rates (0-74) from causes amenable to health care in selected European Union countries, 1980 and 1998

Differences among women are less pronounced; but again, in 1998, amenable mortality was highest in Portugal (96.9/100,000) and lowest in Sweden (51.9/100,000) (not shown).

The findings support the notion that improvements in access to effective health care had a measurable impact in many countries during the 1980s and 1990s. However, the scope for improvement in amenable deaths was greatest in those countries where initial rates were highest. As a consequence, the extent of variation between countries will inevitably decrease as rates fall to low levels in all countries. It therefore seems that, in the twenty-first century, the ability to compare health system performance in industrialized countries using mortality data at the aggregate level may be limited.

Focused analytical studies

A third approach is the focused analytic study that takes a single issue and asks questions such as what are the strengths and weaknesses of different ways of funding health care systems? What are the lessons from experience of hospital reform? What are the challenges in regulating the medical profession?

There are many examples of this approach, such as the studies undertaken by the European Observatory on Health Systems and Policies on topics such as funding health care (Mossialos *et al.* 2002), the role of hospitals (McKee and Healy 2002), and regulating entrepreneurial behaviour in health systems (Saltman *et al.* 2002). This approach is based on the belief that benefits can be obtained by learning from best practice within health systems, in particular by looking at examples from other countries, while recognizing the importance of national specificities and path dependency noted earlier. It builds on a rich literature on lesson drawing, in particular the pioneering work by Rose (1993). This involves a series of questions. First, what policies, that are *already in operation* would work in the exporting setting, drawing on existing evaluative research? Note the emphasis on 'already in operation'. Too frequently, lessons are drawn from concepts that have yet to be put into practice, on the basis of beliefs about what they *might* achieve if ever implemented. This was a feature of much of the debate on quasi-markets in health care in the 1990s (Le Grand and Bartlett 1993). Second, what are the contextual factors that are necessary for it to work in that setting? Third, do those factors exist, to an extent sufficient for policy transfer to take place, in the setting into which the policy is being imported, and in what ways does the policy need to be modified? Fourth, once imported, does the policy work as intended, again introducing the need for evaluative research and thus closing the circle.

Lesson drawing can take a variety of forms, from direct copying, through adaptation, creation of hybrids and acting as a source of inspiration. However, it is also common for policies developed elsewhere not to be transferred but instead used as *post hoc* validation for decisions already taken (Bennett 1991). Clearly lesson drawing depends on both a detailed understanding of the policies being compared and a detailed understanding of the contextual factors that determine whether a policy works in different circumstances. A soft systems approach may be helpful in making explicit both of these elements. Leichter has proposed a framework for contextual analysis, incorporating situational (transient, impermanent, or idiosyncratic conditions or events that impact on policy-making), structural (relatively unchanging elements of the society and polity), cultural (the value system within society) and external (events, structures, and values that exist outside the boundaries of a political system, but that influence decisions in the system) factors (Leichter 1979). However, this remains a relatively under-researched area of comparative health systems research, not least because of its complexity and the need to draw on many areas of knowledge including comparative government, comparative law, anthropology and economics.

Assessing the performance of health services and health systems

The concern with measuring the performance of health services and health systems is not new. As long ago as the 1860s, Florence Nightingale was pioneering the systematic collection and reporting of hospital performance data. In the 1910s, Ernest Codman promoted the need for collection and public release of surgical outcome data. He recorded diagnostic and treatment errors and linked these to outcomes to improve services (Neuhauser 2002). However, it is only recently that

Box 2.7 Why study health system performance?

- *Accelerating advances in medical technology*
Potential for new interventions and methods for delivering and organizing health care, with consequences for health care expenditures, resulting in the need to ensure that innovations – of whatever sort – promote system objectives and avoid adverse side effects (Smith 2002a).
- *Evidence on variation in the use of services often with little variation in health outcomes*
Evidence of overuse, underuse and misuse of services; increase in available technologies and patient expectations (overuse); limited access, trust, fragmentation of services (underuse); and system failures (misuse) (Becher and Chassin 2001).
- *Growing uncertainty about the actual quality and outcomes of health services*
UK: Bristol Inquiry into the deaths of children undergoing heart surgery at the Bristol Royal Infirmary has highlighted the failings of existing internal quality assurance and professional regulation systems (Smith 1998).
USA: According to estimates by the Institute of Medicine at least 44,000 avoidable deaths are caused by medical errors in hospitals (2 per cent of all deaths), which is more than the number of lives lost by motor vehicle accidents per year (Kohn et al. 1999).
- *Rising public expectations*
Demand for 'value for money' as taxpayers or insured; increasing access to information, facilitated by the new electronic media such as the internet; increasing awareness of medical errors, and preparedness to challenge professional authority, to litigate and advocate.
- *New public management culture in publicly-funded health care systems*
Demand for public accountability of public services including financial transparency, creation of market and quasi-market mechanisms such as the introduction of competition between providers or, as in the UK in the early 1990s, the introduction of the internal market.

performance assessment has been put more firmly on the national and international agenda (see Box 2.7).

This section examines some of the main conceptual and methodological issues that underlie our understanding of performance initiatives in the field of health (Wait and Nolte in press). Most often these involve the use of performance indicators or, more broadly, measures that capture a variety of health and health-system related trends and factors. The terminology can, however, sometimes be confusing; Box 2.8 gives an overview of the most common definitions used in this field.

This terminology is common to assessments of individual clinical or organizational interventions and to health systems and their major sub-systems. Depending on the scope of the assessment, it may focus on the measurement of inputs (structure), process or outcomes, or combinations thereof, often subsumed under the heading 'performance indicator' (Wait and Nolte in press). However, when system-wide assessments are undertaken, they are often based on somewhat crude measures, for example health care expenditure, health care resources such as the number of

Box 2.8 Some definitions used in assessing health care performance

• Indicator	'Measurement tool used to monitor and evaluate the quality of important government, management, clinical, and support functions' (Joint Commission on Accreditation of Healthcare Organizations 1990).
• Performance indicator	A 'quantitative measure of quality' (Ibrahim 2001).
• Quality	'[T]he degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge' (Hurtado et al. 2001).
• Performance	Multidimensional concept that 'along with efficiency [incorporates] dimensions of quality (in the sense of safety, effectiveness of care, quality of services rendered and quality as perceived by patients and the people around them) and equity' (Girard and Mirvielle 2002).

physicians, nurses, facilities, etc. or summary measures of population health such as mortality or life expectancy (Hurst and Jee-Hughes 2000). Among the more advanced examples are, at the national level, the National Performance Assessment Framework (PAF) in the UK (Smee 2002; Smith 2002b) and, at the international level, the previously mentioned HSPAF developed by WHO (2000).

Given the broad definition of performance, the scope of performance indicators can potentially be enormous, ranging from assessing national health systems down to patient experience with an individual provider (Wait and Nolte in press). Assessments may thus involve indicators at the *micro level* describing the primary process of patient care, such as assessing or comparing the performance of individual surgeons. For example, the New York State Department of Health has operated a programme since 1989 that collects and makes public data on risk-adjusted mortality following coronary artery bypass surgery by hospital and by surgeon (see Box 2.9). In England, following the Bristol Inquiry in 2001, there have also been plans to publish performance data by individual surgeons and several specialist associations undertake analysis of especially collected data sets as a method of comparative audit. Also in the UK, an independent company, Dr Foster, published data on the mortality experience of hospitals, in collaboration with *The Times* newspaper, as well as a guide to individual specialists (Dr Foster 2004), although in this case looking only at measures of process such as waiting times.

The next level of assessment involves looking at the *meso, or organisational level*, for example, assessing the performance of a primary care team or a hospital. Examples include the public reporting systems in New York and several other American states (Hannan et al. 1997). Within the UK PAF, performance indicators of individual hospitals have been published since 2000. In 2001, this was complemented by a performance rating system in which the progress of NHS Hospital Trusts (health care delivery organizations) in England were assessed against nine key targets, 28 performance indicators and judgements of the Commission for Health Improvement (a health service inspectorate) and, based on this assessment, were awarded

Box 2.9 Measuring the performance of health care providers

Some areas in the USA have been publishing data comparing mortality rates according to individual surgeons over the last ten years. One example is New York State where the State Department of Health (DoH) has been collecting data to assess the quality of care provided to patients undergoing coronary artery bypass grafting (CABG) since 1989, using a registry (Hannan *et al.* 1997). Each hospital collects relevant data and forwards them to the DoH on a quarterly basis.

The data are then processed to produce mortality rates for each hospital and surgeon that take into account the severity of each patient's presenting illness and coexisting conditions. In 1990, the DoH published data on crude, expected and risk-adjusted mortality rates and volume of CABG procedures performed at each hospital in New York State. Thus, while surgeon-specific death rates were also calculated, these were not published, the argument being that low volume of operations would lead to substantial variation in mortality rates and therefore be susceptible to misinterpretation. However, upon release of the hospital-specific death rates a newspaper, the *Newsday*, issued a lawsuit against the DoH under the Freedom of Information Law to also gain access to surgeon-specific data. The DoH lost the case and had to publish the data in 1991. Surgeon-specific death rates have now been published on an annual basis since 1992. A similar system is in place in Pennsylvania.

stars, with three stars relating to highest levels of performance down to zero stars, indicating the poorest levels of performance (Department of Health 2002). Efforts to evaluate performance in the ambulatory sector have been hindered mostly by limited availability of appropriate information; however, in the UK performance ratings have now been extended to include Primary Care Trusts, the organizations that purchase health care (Commission for Health Improvement 2004).

Assessments at the *macro level* (regional, national) or the financing and policy context of health care include the UK PAF as mentioned above and other national initiatives seeking to provide a framework to assess the performance of the health care system. A comprehensive evaluation of these national frameworks was recently provided by Arah and colleagues (Arah *et al.* 2003).

Finally, it is important to define the actual objective of (public) reporting of performance indicators. Objectives may include (i) accountability to funders and other stakeholders; (ii) identification of areas of poor performance and centres of excellence; (iii) facilitation of selection and choice of providers by consumer/patients and purchasers of health care; (iv) provider behaviour change; and (v) providing epidemiological and other public health data (Nutley and Smith 1998). Depending on the objective(s) the target audience thus includes the general public, health care providers, purchasers and policy-makers.

Conceptual problems

There are a number of challenges relating to the development, application and reporting of performance indicators; the following section will reflect briefly on some of the major issues including definitions, underlying data and selection of indicators, methodological issues, interpretation of data and unintended consequences.

Definitions

Some of the more fundamental challenges relate to the definitions underlying the process of performance assessment. Thus, in assessing the performance of health systems, be it at the national or international level, one important question is how one defines the health system, as this will define the performance measures being used. For example, as noted earlier, the *World Health Report 2000* used a rather broad definition of a health system that incorporated the importance of intersectoral action to promote health (WHO 2000). However, perhaps inevitably, as a report from an international organization whose constituents are individual countries, it adopted as its basis for comparison the health systems of those countries. This immediately created a problem. In some countries, as already noted, the financing and delivery of health care is the responsibility of a diverse array of organizations that can only loosely be considered to comprise a system. In other countries, such as Afghanistan, Sierra Leone or the Democratic Republic of the Congo, to take only three of the most obvious examples, it was difficult at that time to argue that there was anything in place that resembled a system of government, or at least one whose writ applied beyond the outskirts of the capital city.

Another important challenge relates to the ideological values underpinning any approach to assessing performance. For example, in their assessment of the world's health systems WHO weighted different indicators to reflect their perceived importance in the overall index of performance (WHO 2000). This raises the fundamental question about whose values count. The assessments published in the *World Health Report 2000* used key informants from around the world, but the case has been made that other people, such as those who use the services, should decide which aspects of the health system matter most (Mulligan *et al.* 2000). Some of these issues have now been addressed in subsequent work by WHO, by undertaking large-scale (household) surveys to assess preferences for health system outcomes from the users' point of view (Murray and Evans 2003).

An equally great problem relates to the question of whether and how the measures adopted for assessing performance conform to the underlying definition of the health system. Again the *World Health Report 2000* provides an illustrative example of inconsistencies in this respect (see Box 2.10).

Other examples include the Commonwealth Fund International Working Group on Quality Indicators initiative (CMF QI) and the related OECD Health Care Quality Indicator Project (HCQI) (Nolte *et al.* 2003; Hussey *et al.* 2004). These initiatives aim at the development of a common set of quality indicators for use in cross-national comparisons of health systems. In its first stage the CMF QI initiative adopted a relatively narrow definition of a health system, focusing on the technical quality of health care, or, more specifically, the appropriateness and effectiveness of care (Hussey *et al.* 2004). Yet, the 21 indicators selected to reflect medical care in five countries also included smoking rates. The authors acknowledged that '[t]he health care system does not have perfect control over people's decision to smoke', but, they argued, 'advice and treatment provided by physicians' had been shown to have an impact on smoking cessation (Hussey *et al.* 2004). This line of reasoning seems, however, slightly at odds with the rather narrow objective of evaluating the technical quality of health care.

Selection of indicators and availability of data

Limited data availability and lack of uniformity of data across different settings pose substantial challenges to most initiatives seeking to assess health system performance. In many parts of the world even basic vital statistics are simply not available,

Box 2.10 The attribution of outcomes to activities in the health system

On the basis that a health system should improve health, should respond to the legitimate demands of those it serves, and should prevent families who suffer from illness becoming impoverished as a result, WHO (2000) defined three measures on which to assess health systems: *health attainment*, *responsiveness*, and *fairness of financing*. For the first two, both absolute levels and distribution, as a measure of equity, were assessed, with health attainment assessed as disability adjusted life expectancy (DALE). The overall performance of each health care system was then determined by combining the weighted scores on each measure.

Yet immediately a paradox emerges. It is apparent that many of the determinants of aggregate health lie outside the health care sector. Thus, the measure of health attainment reflects not only those policies and resulting inputs whose primary intent is to improve health but also policies in a wide range of other sectors, such as education, housing and employment, where the production of health is a secondary goal. Yet the other parameters, responsiveness and fairness of financing, relate solely to the delivery of health care services.

Nolte and McKee (2003) have thus examined how health systems perform when attainment can be more directly attributed to health care. Using the concept of 'avoidable mortality' (see Box 2.6), they have calculated standardized death rates from conditions considered amenable to health care for 19 OECD countries (1998) and generated rankings based on the level of amenable mortality. These rankings were then compared with the rankings produced by WHO based on DALE. This exercise showed that rankings changed for most countries that were included in the study, illustrating that the findings of any performance assessment very much depend on the concepts that underlie them.

because of fragmentary population registration systems and even in some industrialized countries significant gaps exist in coverage of some groups, for example native Americans or Australian Aborigines. Where data exist their usefulness may be restricted due to lack of comparability, which poses particular challenges to international comparisons. Thus, in parallel with its reports describing elements of health systems, the OECD (2003) has undertaken pioneering work in assembling an international database of inputs, processes and outcomes of health systems. In doing so, it has identified many weaknesses in the existing data. For example, figures for numbers of health professionals in some countries are based on head counts, taken from professional registers, while in others they are limited to numbers (or in some cases, whole-time equivalents) in employment (and in some cases, only those working in the state sector). Even the question of how much each country spends on health care is often difficult to answer. Most obviously, there is the problem of defining the boundaries of the system. However, even accurate figures for overall expenditure are themselves of limited use and they are frequently expressed in terms of measures of national wealth, such as GNP per capita. Knowledge of this figure is not always easy, especially in less developed countries where the size of the population may be uncertain.

This problem is further highlighted by the *World Health Report 2000* (WHO 2000). Assessing the performance of the health systems of 191 countries required

many heroic assumptions, not least in relation to the virtual absence of data from a majority of the countries involved. In a recent critique, Musgrove showed that only 39 per cent of the indicator values included in the *World Health Report* were based on existing data, the remainder being estimates using regression analyses and other means (Musgrove 2003). Using complex models to generate estimates fails, however, to tackle the underlying problem.

Accurate collection of indicator data relies on the existence of reliable and well-established health information systems. However, most existing systems were originally devised for internal mechanisms of financial control, and their adaptation for purposes of performance assessment may not be straightforward. Problems with minimum data sets, inaccuracies in interpretation of aggregated data, failure to integrate population- and patient-level data and lack of linkage between diagnostic data and outcomes of care are some of the main drawbacks reported in existing health information systems (Shaw and Kalo 2002). With these caveats in mind, the value of performance initiatives can be greatly enhanced if target indicators are selected for their relevance and usefulness as evaluation tools rather than merely on data availability. Indeed, indicators often seem to be selected on the basis of what is available and practical rather than what is meaningful, such as areas that need improvement and require prioritization or health system goals and values (Walshe 2003).

Several groups have presented lists of desirable attributes for performance indicators. According to Pringle and colleagues these should be valid, communicable, effective, reliable, objective, available, contextual, attributable, interpretable, comparable, remediable and repeatable (Pringle *et al.* 2002). The CMF QI selected performance indicators based on (i) feasibility: indicators are already being collected by one or more countries; (ii) scientific soundness: indicators have to be reliable and valid; (iii) interpretability: indicators have to allow a clear conclusion for policy-makers; (iv) actionability: measures can be directly affected by the health care system, and (v) importance: indicator reflects important health conditions in terms of burden of disease, cost of care or priorities of policy-makers (Hussey *et al.* 2004).

Methodological challenges

The methodological challenges to performance assessment or, more generally, evaluation of health systems are manifold and are related to the underlying data, variation in information needs of different users, questions about the actual link between specific inputs and processes of health care and health outcomes, possible time lags between interventions and outcome and the timing of measurements, etc. Also, not all outcomes that are valued by society are measurable – for example, how does one assess reassurance?

One example is emergency readmission to hospital, which is often used as proxy measure of avoidable adverse outcomes after initial admission to hospital, for example in the English NHS performance ratings mentioned above. The use of this indicator is usually justified because a high proportion of emergency readmissions should be preventable if the preceding care is adequate (Leng *et al.* 1999). The appropriateness of this measure, or of readmission to hospital more generally, as a quality or performance indicator has been questioned as other factors unrelated to the quality of hospital care can affect the likelihood of readmission, including patient factors such as severity and chronicity of the underlying condition or levels of co-morbidity, or hospital factors such as validity of administrative data. In addition, variation of (emergency) readmission rates between hospitals may be due to factors such as variation in population structure (ageing population, elderly living alone), falling length of hospital stay, variation among hospitals in case mix and

severity, and issues such as random variation due to small numbers and problems in defining the denominator – again factors not related to the actual quality of care. This is further illustrated by a case study undertaken in Scotland that analysed emergency readmission rates in some detail (Leng *et al.* 1999). It showed that about 17 per cent of the emergency readmissions recorded were unrelated to the initial admission and therefore did not reflect the previous quality of care. The challenge is thus to establish a *causal relationship* between patient outcome and the actual process of care, or, as discussed in Box 2.10, between health outcome at the population level and elements of the health system.

Another issue relates to the use of composite measures. One example is the *World Health Report 2000*, which assessed overall health system performance as a weighted composite measure of health attainment, responsiveness and fair financing in relation to what might be expected given the country's level of economic and educational development (WHO 2000), another is the NHS star rating system described earlier (Department of Health 2002). The use of such measures has been challenged on conceptual and methodological grounds. Thus, referring to the methods employed by WHO, Naylor and colleagues argued that composite indices of health system performance are, at best, of 'dubious precision' for they 'combine uncertain weighing systems, imprecision arising from the potential non-comparability of component measures, and misleading reliability in the form of whole-population averages that mask distributional issues' (Naylor *et al.* 2002). Moreover, presentation of even disaggregated data as means and medians may be misleading since this is likely to conceal fluctuations at various levels within the health system. Naylor *et al.* thus concluded that '[H]ealth systems are extraordinarily complex. In consequence, one must beware of the seductive reductionism of devising a single measure to capture all dimensions of health status, let alone health system performance. A balanced approach with an array of indicators is desirable, as each set of stakeholders will need a different type of information to make better decisions' (2002).

Interpretation and unintended consequences

We have known incurable cases discharged from one hospital, to which the deaths ought to have been accounted and received into another hospital, to die there in a day or two after admission, thereby lowering the mortality rate of the first at the expense of the second.
(Florence Nightingale 1863)

A final question needing to be asked is whether national performance initiatives can contribute to improving the performance of the health care systems they are assessing (see also Box 2.11). Several scholars have expressed concern that the use of indicators has become an end in itself and have urged the evaluation of existing performance indicator systems (Goddard *et al.* 2000; Walshe 2003). Such evaluations are needed to assess the impact and validity of the indicator systems used, their contribution to increasing accountability through the performance management process and their ability to truly reflect the goals and objectives set out by the health care system.

The reporting of performance indicators can have different objectives, such as increasing consumer choice and facilitating change in provider behaviour and, ultimately, improving the performance of the health care system. There are, however, relatively few data to assess whether and how well performance indicator systems achieve any of these objectives. Evidence from the USA suggests that consumers as well as purchasers or payers rarely search out publicly available information and, if they do, do not understand or trust it (Marshall *et al.* 2000). Also,

Box 2.11 Advantages and disadvantages of league tables

League tables are increasingly being used to rank performance in many different sectors including health care and education (Nutley and Smith 1998; Adab *et al.* 2002). League tables are intuitively appealing, especially to politicians who are anxious to know how public funds are being spent. It is seen as a means to reduce a mass of complex information into a format that almost anyone can understand.

Advantages of performance league tables include that they may stimulate competition between providers and may consequently lead to service improvement as providers will adopt best practices. The New York experience has been listed as one example although the findings have been challenged (Chassin 2002). It was also argued that league tables may improve patient or consumer choice, which has been interpreted as a necessity for an efficient market economy. In addition, league tables have the potential to facilitate regulators to monitor and ensure accountability of providers. Regulators can use league table rankings to identify clinicians or hospitals with a high frequency of selected adverse outcomes as a starting point for further enquiry.

On the downside, however, the apparent simplicity of league tables can be quite misleading, and many commentators have drawn attention to the numerous technical problems as discussed in the main text. Also, the aspect of increased consumer choice will only be relevant in health systems where there is indeed a choice of provider and thus largely excludes countries like the UK where patients have relatively little choice when using an individual practitioner or hospital. Finally, while league tables may have the potential to improve quality of services by encouraging providers to put more emphasis on quality of care, they may also have unintended consequences such as distorting priorities by encouraging providers to focus on performance measures *per se* rather than quality improvement ('gaming'). A related problem is that with ranking performance there is an implicit assumption that providers located at the bottom of the table provide a poorer service. This gives way to the development of a culture of naming and shaming in which the blame is often apportioned to individuals.

physicians appear to be rather sceptical about the data and only a small proportion apparently uses them. The US experience thus seems to suggest that 'public disclosure of information about the quality of health care is a weak strategy for ensuring quality' (Schneider and Lieberman 2001). However, there is also evidence suggesting that managers and some providers do use comparative information, with data from the USA showing that hospitals appear to have been most responsive to publicized data with some evidence pointing towards improvements in care where public reporting occurred (Marshall *et al.* 2000; Chassin 2002). Based on this and experience elsewhere, Leatherman (2002: 329) thus concluded that '[t]he state of the art of performance measurement and reporting has made dramatic advances in the past decade but it is still deficient to support widespread diffusion, predictable systematic application, and routinely fair and accurate assessments'.

Conclusions

The quest for a means of evaluating an entire health system is far from simple. Health systems are intrinsically complex entities, with flexible boundaries, the definitions of which depend on the question being asked. While the main goals of a health system can easily be defined, it is more difficult to identify a way of assessing whether these goals are being achieved and the extent to which apparent progress can be attributed to the health system or to other factors. It is important not to overlook the many intermediate or subsidiary goals of a health system, as policies designed to achieve one goal may impact adversely on progress towards another. Health systems, like all human systems, are adaptable, so that the impact of an intervention designed to bring about change can be difficult to predict. They are also contextually bounded, so that something that works in one country may not work in the same way in another.

If there are key messages that can be taken from this chapter they are, first, that in evaluating some aspect of a health care system one must begin by defining the question being asked as precisely as possible and, second, that the evaluation must be informed by the context within which each system exists.

There is an inevitable tension between the simple answers often sought by politicians, such as whether the health system in country A is better than that in country B, and the messy complexity that gives rise to the analysts answer that 'it depends on what you mean'. Instead, by cataloguing the many challenges that exist, this chapter may act as a stimulus for both groups to come together in a constructive dialogue that will enable the former to define their questions more precisely and the latter to develop ways in which these improved questions can be answered.

A rectangular box with a stippled background and a black border. The text 'Key points' is written in a bold, sans-serif font. A vertical line extends upwards from the top center of the box, and another vertical line extends downwards from the bottom center of the box.

Key points

- There are different definitions of what a health system is.
- The approach taken to describing and analysing health systems depends critically on the question being asked.
- In evaluating health systems, the question being asked must be defined as precisely as possible.
- Thus any definition used must be explicit and the means of evaluating the system must be congruent with the definition.
- The evaluation must be informed by the context within which each system exists.
- There has been a move away from the evaluation of the system as a whole towards the evaluation of sub-systems.
- Approaches to health system comparison fall broadly into one of three main groups: descriptive studies, quantitative approaches and focused analytical studies.
- Challenges to the development, application and reporting of performance indicators include: the definitions underlying the process of performance assessment; the ideological values underpinning any approach to this; limited data availability and lack of uniformity of data across different settings; and methodological issues, including the validity of proxy composite measures.

Acknowledgement

The work of Ellen Nolte and Suzanne Wait on international benchmarking in health is supported by fellowships from the Nuffield Trust.

References

- Abel-Smith, B. (1963) Paying for health services: a study of the costs and sources of finance in six countries, *Public Health Papers*, 17: 1–86.
- Abel-Smith, B. (1967) An international study of health expenditure and its relevance for health planning, *Public Health Papers*, 32: 1–127.
- Adab, P., Rouse, A.M., Mohammed, M.A. and Marshall, T. (2002) Performance league tables: the NHS deserves better, *British Medical Journal*, 324: 95–8.
- Anderson, O.W. (1963) Medical care: its social and organizational aspects. Health-services systems in the United States and other countries – critical comparisons, *New England Journal of Medicine*, 269: 839–43.
- Anderson, O.W. (1972) *Health Care: Can there be Equity? The United States, Sweden and England*. New York: Wiley.
- Arah, O.A., Klazinga, N.S., Delnoij, D.M.J., Ten Asbroek, A.H.A. and Custers, T. (2003) Conceptual framework for health systems performance: a quest for effectiveness, quality and improvement, *International Journal of Quality in Health Care*, 15: 377–98.
- Becher, E.C. and Chassin, M.R. (2001) Improving the quality of health care: who will lead? *Health Affairs*, 20: 164–79.
- Bennett, C.J. (1991) How states use foreign evidence, *Journal of Public Policy*, 11: 31–54.
- British Broadcasting Corporation (2000) French health care 'best in world', <http://news.bbc.co.uk/1/hi/health/799444.stm> (accessed 8 May 2004).
- Buck, D., Eastwood, A. and Smith, P.C. (1999) Can we measure the social importance of health care? *International Journal of Technology Assessment in Health Care*, 15: 89–107.
- Charlton, J.R.H., Hartley, R.M., Silver, R. and Holland, W.W. (1983) Geographical variation in mortality from conditions amenable to medical intervention in England and Wales, *Lancet*, i: 691–6.
- Chassin, M.R. (2002) Achieving and sustaining improved quality: lessons from New York State and cardiac surgery, *Health Affairs*, 21: 40–51.
- Checkland, P. (1981) *Systems Thinking, Systems Practice*. Chichester: Wiley.
- Commission for Health Improvement (2004) <http://www.chi.nhs.uk/ratings/> (accessed 2 May 2004).
- Department of Health (2002) *NHS Performance Ratings*. London: Department of Health.
- Donabedian, A. (1966) Evaluating the quality of medical care, *Milbank Quarterly*, 44: 166–203.
- Donabedian, A. (1980) *The Definition of Quality and Approaches to its Management*. Ann Arbor, MI: Health Administration Press.
- Donabedian, A. (1988) The quality of care: how can it be assessed? *JAMA*, 260: 1743–8.
- Dr Foster (2004) <http://www.drfooster.co.uk/home.asp> (accessed 2 May 2004).
- Elling, R.H. (1980) *Cross-national Study of Health Systems: Concepts, Methods and Data Sources. Health Affairs Information Guide Series 2*. Detroit, MI: Gale Research Company.
- Elling, R.H. (1994) Theory and method for the cross-national study of health systems, *International Journal of Health Services*, 24: 285–309.
- European Observatory on Health Systems and Policies (2004) <http://www.who.dk/observatory> (accessed 30 April 2004).
- Evans, R.G. (1981) Incomplete vertical integration: the distinctive structure of the health-care industry, in J. Van Der Gaag and M. Perlman (eds) *Health, Economics and Health Economics*. Amsterdam: North Holland.
- Field, M.G. (1973) The concept of the 'health system' at the macrosociological level, *Social Science & Medicine*, 7: 763–85.
- Field, M.G. (1978) *Comparative Health Systems: Differentiation and Convergence*. Washington: National Center for Health Services Research.

- Frenk, J. (1994) Dimensions of health system reform, *Health Policy*, 27: 19–34.
- Frenk, J. and Donabedian, A. (1987) State intervention in medical care: types, trends and variables, *Health Policy and Planning*, 2: 17–31.
- Girard, J.-F and Minvielle, E. (2002) Measuring up: lessons and potential, in *Measuring Up: Improving Health System Performance in OECD Countries*. Paris: OECD.
- Goddard, M., Mannion, R. and Smith, P. (2000) Enhancing performance in health care: a theoretical perspective on agency and the role of information, *Health Economics*, 9: 95–107.
- Goldman, F. (1946) Foreign programs of medical care and their lessons, *New England Journal of Medicine*, 234: 156.
- Green, D.G., Irvine, I., Mckee, M., Dixon, A. and Mossialos, E. (2002) For and against: social insurance – the right way forward for health care in the United Kingdom? *British Medical Journal*, 325: 488–90.
- Hannan, E.L., Stone, C.C., Biddle, T.L. and Debuono, B.A. (1997) Public release of cardiac surgery outcomes data in New York: what do New York state cardiologists think of it? *American Heart Journal*, 134: 1120–8.
- Holland, W.W. (1988) *European Community Atlas of 'Avoidable Death'*. Oxford: Oxford University Press.
- Hurst, J. and Jee-Hughes, M. (2000) *Performance Measurements and Performance Management in OECD Health Systems*. Paris: OECD.
- Hurtado, M.P., Swift, E.K. and Corrigan, J.M. (2001) *Envisioning the National Health Care Quality Report*. Washington: National Academy Press.
- Hussey, P.S., Anderson, G.F., Osborn, R., Feek, C., McLaughlin, V., Millar, J. and Epstein, A. (2004) How does the quality of care compare in five countries? *Health Affairs*, 23: 89–99.
- Ibrahim, J.E. (2001) Performance indicators from all perspectives, *International Journal of Quality in Health Care*, 13: 431–2.
- International Network for Health Policy and Reform (2004) <http://www.health-policy-monitor.org/index.jsp> (accessed 30 April 2004).
- Joint Commission on Accreditation of Healthcare Organizations (1990) *Primer on Indicators Development and Application*. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations.
- Kanavos, P. and Mossialos, E. (1990) International comparisons of expenditures: what we know and what we do not know, *Journal of Health Services Research and Policy*, 4: 122–6.
- Klein, R. (1997) Learning from others: shall the last be first? *Journal of Health Politics, Policy and Law*, 22: 1267–78.
- Kohn, L.T., Corrigan, J.M. and Donaldson, M.S. (1999) *To Err is Human: Building a Safer Health System*. Washington, DC: National Academic Press.
- Last, J.M. (2001) *A Dictionary of Epidemiology*. Oxford: Oxford University Press.
- Le Grand, J. and Bartlett, W. (1993) *Quasi-Markets and Social Policy*. London: Palgrave Macmillan.
- Leatherman, S. (2002) Applying performance indicators to health system improvement, in *Measuring Up: Improving Health System Performance in OECD Countries*. Paris: OECD.
- Leichter, H.M. (1979) *A Comparative Approach to Policy Analysis: Health Care Policy in Four Nations*. Cambridge: Cambridge University Press.
- Leng, G.C., Walsh, D., Fowkes, F.G.R. and Swainson, C.P. (1999) Is the emergency readmission rates a valid outcome indicator? *Quality in Health Care*, 8: 234–8.
- Lewin, R. (1992) *Complexity: Life at the Edge of Chaos*. Chicago: University of Chicago Press.
- Long, M.J. (1994) *The Medical Care System: A Conceptual Model*. Ann Arbor, MI: AUPHA Press.
- Macinko, J., Starfield, B. and Shi, L. (2003) The contribution of primary care systems to health outcomes within Organization for Economic Cooperation and Development (OECD) countries, 1970–1998, *Health Services Research*, 38: 831–65.
- Marshall, M.N., Shekelle, P.G., Leatherman, S. and Brook, R.H. (2000) The public release of performance data: what do we expect to gain? A review of the evidence, *JAMA*, 283: 1866–74.
- Maxwell, R. (1974) *Health Care, the Growing Dilemma: Needs Versus Resources in Western Europe, the US and the USSR*. New York: McKinsey & Co.
- McKee, M. (1995) 2020 vision, *Journal of Public Health Medicine*, 17: 127–31.
- McKee, M. and Figueras, J. (1997) Comparing health care systems: how do we know if we can learn from others? *Journal of Health Services Research and Policy*, 2: 122–5.

- McKee, M. and Healy, J. (2002) *Hospitals in a Changing Europe*. Buckingham: Open University Press.
- McPake, B., Kumaranyake, L. and Normand, C. (2002) *Health Economics: An International Perspective*. London: Routledge.
- Mechanic, D. (1975) Ideology, medical technology, and health care organization in modern nations, *American Journal of Public Health*, 65: 241–7.
- Mills, A. and Ranson, M. (2001) The design of health systems, in M. Merson, R. Black and A. Mills (eds) *International Public Health, Diseases, Programs, Systems, and Policies*. Gaithersburg, MD: Aspen Publications.
- Milne, R. and Molana, H. (1991) On the effect of income and relative price on demand for health care: EC evidence, *Applied Economics*, 23: 1221–6.
- Mossialos, E., Dixon, A., Figueras, J. and Kutzin, J. (2002) *Funding Health Care: Options for Europe*. Buckingham: Open University Press.
- Mountin, J.W. and Perrott, G.S. (1947) Health insurance programs and plans of western Europe: summary of observations, *Public Health Reports*, 62: 369–99.
- Mulligan, J., Appleby, J. and Harrison, A. (2000) Measuring the performance of health systems, *British Medical Journal*, 321: 191–2.
- Murray, C.J.L. and Evans, D.B. (2003) *Health Systems Performance Assessment: Debates, Methods and Empiricism*. Geneva: WHO.
- Murray, C.J.L. and Frenk, J. (2000) A framework for assessing the performance of health systems, *Bulletin of the World Health Organization*, 78: 717–31.
- Musgrove, P. (2003) Judging health systems: reflections on WHO's methods, *Lancet*, 361: 1817–20.
- Naylor, C.D., Iron, K. and Handa, K. (2002) Measuring health system performance: problems and opportunities in the era of assessment and accountability, in *Measuring Up: Improving Health System Performance in OECD Countries*. Paris: OECD.
- Neuhauser, D. (2002) Ernest Amory Codman MD, *Quality and Safety in Health Care*, 11: 104–5.
- Nolte, E. and McKee, M. (2003) Measuring the health of the nations: how much is attributable to health care? An analysis of mortality amenable to medical care, *British Medical Journal*, 327: 1129–32.
- Nolte, E. and McKee, M. (2004) *Does Healthcare Save Lives? Avoidable Mortality Revisited*. London: The Nuffield Trust.
- Nolte, E., Wait, S., Bain, C. and McKee, M. (2003) What can be measured or what is important? A critical analysis of international comparisons of health system performance. Poster presented at 5th International Conference on the Scientific Basis of Health Services, Washington, 20–3 September.
- Nutley, S. and Smith, P.C. (1998) League tables for performance improvement in health care, *Journal of Health Services Research and Policy*, 3: 50–7.
- OECD (1985) *Measuring Health Care: 1960–1983 Expenditure, Costs and Performance*. Paris: OECD.
- OECD (1992) *The Reform of Health Care Systems: A Comparative Analysis of Seven OECD Countries*. Paris: OECD.
- OECD (2002) *Measuring Up: Improving Health System Performance in OECD Countries*. Paris: OECD.
- OECD (2003) *OECD Health Data 2003: A Comparative Analysis of 30 OECD Countries*. Paris: OECD.
- Or, Z. (2000) Determinants of health outcomes in industrialised countries: a pooled, cross-country, time-series analysis, *OECD Economic Studies*, 30: 53–77.
- Or, Z. (2001) *Exploring the Effects of Health Care on Mortality Across OECD Countries*. Paris: OECD.
- Parkin, D., Mcguire, A. and Yule, B. (1987) Aggregate health care expenditures and national income: is health care a luxury good? *Journal of Health Economics*, 6: 109–27.
- Plochg, T. and Klazinga, N.S. (2002) Community-based integrated care: myth or must? *International Journal for Quality in Health Care*, 14: 91–101.
- Pollock, A.M. and Price, D. (2003) The public health implications of world trade negotiations on the general agreement on trade in services and public services, *Lancet*, 362: 1072–5.
- Pringle, M., Wilson, T. and Grol, R. (2002) Measuring 'goodness' in individuals and health-care systems, *British Medical Journal*, 325: 704–7.

- Raffel, M.W. (1984) *Comparative Health Systems – Descriptive Analyses of Fourteen National Health Systems*. University Park, PA: Pennsylvania State University Press.
- Reinhardt, U.E., Hussey, P.S. and Anderson, G.F. (2002) Cross-national comparisons of health systems using OECD data, 1999, *Health Affairs*, 21: 169–81.
- Relman, A.S. (1988) Assessment and accountability: the third revolution in medical care, *New England Journal of Medicine*, 319: 1220–2.
- Rodwin, V.G. (1984) *The Health Planning Predicament: France, Quebec, England, and the United States*. Berkeley, CA: University of California Press.
- Roemer, M.I. (1960) Health departments and medical care – a world scanning, *American Journal of Public Health*, 50: 154.
- Roemer, M.I. (1969) *The Organization of Medical Care Under Social Security*. Geneva: International Labour Office.
- Roemer, M.I. (1977) *Comparative National Policies on Health Care*. New York: Marcel Dekker.
- Roemer, M.I. (1991) *National Health Systems of the World*. New York: Oxford University Press.
- Rose, R. (1993) *Lesson Drawing in Public Policy*. Chatham, NJ: Chatham House.
- Rutstein, D.D., Berenberg, W., Chalmers, T.C., Child, C.G., Fishman, A.P. and Perrin, E.B. (1976) Measuring the quality of medical care, *New England Journal of Medicine*, 294: 582–8.
- Saltman, R.B., Busse, R. and Mossialos, E. (2002) *Regulating Entrepreneurial Behaviour in European Health Care Systems*. Buckingham: Open University Press.
- Schieber, G.J. (1987) *Financing and Delivering Health Care: A Comparative Analysis of OECD Countries*. Paris: OECD.
- Schneider, E.C. and Lieberman, T. (2001) Publicly disclosed information about the quality of health care: response of the US public, *Quality in Health Care*, 10: 96–103.
- Shaw, C. and Kalo, I. (2002) *A Background for National Quality Policies in Health Systems*. Copenhagen: WHO Regional Office for Europe.
- Sheaff, R. (1998) Towards a global theory of health systems: Milton Roemer's *National Health Systems of the World*, *Health Care Analysis*, 6: 150–70.
- Sigerist, H.E. (1943) From Bismarck to Beveridge: developments and trends in social security legislation, *Bulletin of the History of Medicine*, 8: 365–88.
- Smee, C.H. (2002) Improving value for money in the United Kingdom National Health Service: performance measurement and improvement in a centralised system, in *Measuring Up: Improving Health System Performance in OECD Countries*. Paris: OECD.
- Smith, P. (2002a) Editor's preface, in *Measuring Up: Improving Health System Performance in OECD Countries*. Paris: OECD.
- Smith, P. (2002b) Performance management in British health care: will it deliver? *Health Affairs*, 21: 103–32.
- Smith, R. (1998) Regulation of doctors and the Bristol Inquiry, *British Medical Journal*, 317: 1539–40.
- Terris, M. (1978) The three world systems of medical care: trends and prospects, *American Journal of Public Health*, 65: 1125–31.
- Thompson, C.R. and McKee, M. (2004) Financing and planning of public hospitals in the European Union, *Health Policy*, 67: 281–91.
- Wait, S. and Nolte, E. (in press) Benchmarking health systems: trends, conceptual issues and future perspectives, *Benchmarking: An International Journal*.
- Walshe, K. (2003) International comparisons of the quality of health care; what do they tell us? *Quality and Safety in Health Care*, 12: 4–5.
- Weber, M. (1950) *On Methodology of Social Sciences*. New York: Macmillan.
- Weinerman, E.R. (1971) Research on comparative health service systems, *Medical Care*, 9: 272–90.
- WHO (World Health Organization) (1948) Preamble to the *Constitution of the World Health Organization* as adopted by The International Health Conference, New York 19–22 June 1946. Geneva: World Health Organization.
- WHO (World Health Organization) (2000) *The World Health Report 2000: Health Systems, Improving Performance*. Geneva: World Health Organization.
- WHO (World Health Organization) Regional Office For Europe (2004) *Health for All Database*. Copenhagen: WHO Regional Office for Europe.
- Wilson, T. and Holt, T. (2001) Complexity and clinical care, *British Medical Journal*, 323: 685–88.