© OECD, 2003.

© Software: 1987-1996, Acrobat is a trademark of ADOBE.

All rights reserved. OECD grants you the right to use one copy of this Program for your personal use only. Unauthorised reproduction, lending, hiring, transmission or distribution of any data or software is prohibited. You must treat the Program and associated materials and any elements thereof like any other copyrighted material.

All requests should be made to:

Head of Publications Service,
OECD Publications Service,
2, rue André-Pascal,
75775 Paris Cedex 16, France.

© OCDE, 2003.

© Logiciel, 1987-1996, Acrobat, marque déposée d'ADOBE.

Tous droits du producteur et du propriétaire de ce produit sont réservés. L'OCDE autorise la reproduction d'un seul exemplaire de ce programme pour usage personnel et non commercial uniquement. Sauf autorisation, la duplication, la location, le prêt, l'utilisation de ce produit pour exécution publique sont interdits. Ce programme, les données y afférantes et d'autres éléments doivent donc être traités comme toute autre documentation sur laquelle s'exerce la protection par le droit d'auteur.

Les demandes sont à adresser au :

Chef du Service des Publications,
Service des Publications de l'OCDE,
2, rue André-Pascal,
75775 Paris Cedex 16, France.

Chapter 12

MEASURING THE QUALITY OF HOSPITAL CARE: THE STATE OF THE ART

WHAT INFORMATION SHOULD BE MADE AVAILABLE TO THE PUBLIC?

by

Gérard de Pouvourville and Étienne Minvielle*

Abstract

The quality of hospital care has become a focal point in the regulation of health systems. This can be put down to several factors: the funding crisis, greater public demands on health care professionals owing in part to the wider availability of health-related information and, finally, the development of new data processing technologies. This combination of factors has caused the medical profession's monopoly to be called into question and has prompted calls for the development of external schemes to monitor, and even control, service provision.

The effect of these demands has been the emergence of a new field of health research focussing on the measurement of health care quality, and in particular outcomes, and initiatives by payers or government agencies aimed at making outcome measurement an integral feature of hospital resource allocation. At the same time, those same players and the popular press have been providing the public with "league tables" comparing hospitals on the basis of selected indicators. The nature and quality of these initiatives vary substantially across OECD countries. The United States, with its competition-led health care system, has been at the forefront in both methodological research and the introduction of external control schemes. In the European and other OECD Member countries, the publication of similar research has taken longer, partly for political reasons but also because of delays in computerisation.

The aim of this paper is to gain insights from experiments conducted in a selection of countries. It addresses two points in greater depth. First, it reviews the state of the art regarding methods of measuring health care quality and outcomes, the main emphasis being on the measurement of hospital care quality in terms of either clinical outcomes or health status. Second, it analyses the use made of these methods by external institutions – payers, government agencies and the popular press – looking at why and how they are used, the problems encountered and, finally, how they are currently assessed and what recommendations can be derived from them.

Introduction

The cost-containing constraints that have been placed on health-care systems in developed countries have led public and private financial sources alike to pay greater attention to unit costs in the delivery of health services. At the same time, however, concern over the quality of care has been growing, with an insistence that efforts to reduce or moderate cost trends should not result in a lower quality of care. On this point, the various stakeholders have adopted different approaches, which may be classified

* Centre for Health Economics and Administration Research, INSERM/CNRS, 80 rue du Général Leclerc, 94276 Le Kremlin Bicêtre Cedex France, *gdepouvo@kb.inserm.fr*

© OECD 2002

into four categories: the professional approach, that of rationalizing practices; the managerial approach, drawing upon experience with managing continuous quality improvement in industry; a competition-based approach, seeking to force health-care providers to publish indicators of the quality of their services; and finally a public policy approach, with the authorities playing a key role in the publication of those indicators either in the name of "accountability" or as an effort to overcome the asymmetry of information between potential users and providers of health services. The latter two approaches lead to the same result – the publication of quality indicators as a basis for evaluating and comparing service performance – but the *modus operandi* is distinct in each case: in one case, it is those who pay for care (most often the insurers) who are taking advantage of competition to demand information from care providers, while in the second case intervention is accomplished through regulation.

These approaches have been implemented at varying paces and in different ways from one country to the next. The United States has had a head start over other OECD countries. For example, the federal Health Care Finance Agency (HCFA), the funding manager for Medicare, has been publishing adjusted mortality rates for Diagnosis-Related Groups since 1986. The Joint Commission on the Accreditation of Health Care Organizations (JCAHO) publishes a guide to the different ways of measuring hospital performance (JCAHO, 2001). In 1991, a number of private businesses and non-profit organisations launched the National Committee for Quality Assurance (NCQA) to provide information on the various health insurance plans available to Americans. This progress can be explained by four characteristics of the United States: the federal and state governments are required to report regularly to their legislatures on their actions; there is a strong belief in competition as the principal regulator of economic activities; consequently , there are many players in the health insurance system; and finally, the health-care system is under constant pressure from the courts in the form of malpractice suits. The United Kingdom has had in place since 1989 a policy of fostering inter-hospital competition, including the publication of regular comparative data on hospital performance. England has limited itself to publishing indicators that reflect the quality of service provided (for example waiting lists) (NHS, 1995) rather than clinical outcomes, and it is only in the last year that the National Health Service has published hospital mortality data. In 2001, however, a private initiative known as "Dr. Foster"[1] has begun posting comparative data on public and private hospitals on an Internet site: these data have sparked lively debate, given the method used to adjust mortality rates. Outcome and process indicators are regularly published in Scotland (Shaw, 1987). Canada publishes an annual survey that compares the quality of services available by geographic zone (Marshall, 2001). It is prepared by the Canadian Institute for Health Information (CIHI, 2001). This annual survey is complementary to an initiative of the Province of Ontario, which has published "Hospital Report Cards" on 95 provincial institutions. In France, it was the media that took the initiative, with publication in 1999 of a hospital quality ranking (*Science et Avenir*, 1999). This list has since been regularly reproduced in other magazines (*Figaro Magazine*, *Le Point*). Since then, the Ministry of Health has prepared a French model for publishing comparative data on hospital performance (DREES, 1999, 2001), as has the National Health Care Accreditation and Evaluation Agency (ANAES) (Box 1).

This paper attempts to summarize these experiments, focusing on the measurement of hospital performance. The approach excludes the many US studies that evaluate not only health-care providers taken individually but also the health-care plans marketed by "managed care organisations". In the latter case, an evaluation must cover all the services offered by a plan.[2] The paper also limits itself to measuring the clinical outcomes of health care, *i.e.* performance in terms of reduced morbidity, mortality avoided, improved biological and clinical parameters for patients. This excludes the impact of health care on patients' quality of life, which is still at a very experimental stage (Cooper *et al.*, 2001). It also excludes those dimensions of service quality that relate primarily to the patient's physical and psychological comfort, which can be measured either by specific indicators (length of waiting lists) or by patient experience surveys (Coulter and Cleary, 2001).[3] Finally, the paper does not take into account studies on "total quality improvement" management methods, the objective of which is to ensure that an institution maintains permanent control over factors that determine the quality of care.

© OECD 2002

Box 1. **The Ontario Hospital Report 2001: care indicators**

This comparison of hospital care is based on a selection of ten medical conditions, for which four types of indicators are measured. The conditions are the following:

- Myocardial infarction
- Stroke
- Heart failure
- Community-acquired pneumonia
- Asthma
- Intestinal haemorrhage
- Cholecystectomy
- Hysterectomy
- Prostatectomy
- Carpal tunnel syndrome

The indicators are the following: the number of complications in comparison to the number of cases, readmissions, length of stay, and access to technology or day-hospital care.

The first part of the paper presents what we may call the "technology" of measuring hospital performance. The first segment reviews the principal characteristics of hospital care services, while the second lists their consequences for quality measurement.

The second part of the paper presents the indicators most widely used in various countries, and discusses the limits of some of these indicators.

The third part examines the ways this technology is used and attempts to identify its impacts, on the basis of published information. It looks first at the objectives of those who advocate the production and publication of hospital performance measures, and then examines the impact of these efforts from three points of view: that of the patient, that of the care purchaser (public or private insurers), and that of the authorities.

The final part draws some lessons from the analysis and offers some recommendations for action.

1. Hospital performance measurement technology

1.1. *The characteristics of hospital care*

A *hospital is, first of all, a multi-service provider.* Using the most common tools for measuring hospital workload, which are systems of the Diagnosis-Related Groups type for classifying hospital stays, a hospital may "produce" between 600 and 1 500 different types of hospitalisation, depending on the degree of refinement of the classification used. This very variety poses a problem, particularly if one is seeking to measure not only workload but also productivity and quality of service. In fact, it requires rigorous measures for each "product". It also means that a hospital can perform well in some activities, and less so in others. At the present time, however, there are no data covering all of a hospital's potential activities.

The first solution is to select a small number of activities, generally those that are most in the media spotlight (heart surgery or obstetrics), those that are frequently performed but are not very serious (treatments that pose no risk to patients), those of an elective nature (where there is a high risk of unnecessary treatment), or those selected on the basis of the target population (hip fractures among the elderly) or the availability of proven databases (resuscitation units). The second solution is to look for indicators that will reflect the overall results of care, such as casemix-adjusted mortality rates. Choosing the first method means that we will have no comprehensive comparison of hospital performance. For patients suffering from a particular condition not covered by the selected indicators, published information will be of

© OECD 2002

no interest. To remedy this problem, Eddy (1998) suggests a systematic shift of target pathologies from one year to the next. Mortality rates pose some specific problems, which will be addressed below.

The second characteristic of hospital care is that it represents a complex personal service. The hospital product involves not merely providing a bed but taking charge of an individual, his illness, the dependency induced by that illness, and its treatment. This responsibility involves many inputs: the combination of inputs selected must not only be based on the best available knowledge, but must be delivered efficiently and must be appropriate to each case. This is what Minvielle (1996) calls "managing patient singularity on a large scale". In a service relationship, the patient himself is an input to the process, as a function of his state of health at admission and his genetic inheritance, on one hand, and his degree of cooperativeness under treatment, on the other hand. Finally, the quality of the interpersonal relationship between the patient, his family and the hospital team also contributes to the outcome of the process. For this reason, there is often some confusion about the notion of care quality, *i.e.* between the quality of care considered as an attribute of the outcome of the process (curing a disease, overcoming a handicap, etc.) and the quality of care considered as an attribute of the process itself. It is usual, in the specialized literature on health-care evaluation, to distinguish between two notions of quality: outcome measures and process measures. Among the outcome measures, we may also distinguish between intermediary outcomes, such as controlling high blood pressure or cholesterol levels, and final outcomes, such as cure, remission or death. Finally, the fact that the patient may be both the beneficiary of the process and an input resource to it means that, in comparing different hospitals, their clientele characteristics must be taken into account for their impact on outcomes.

From a theoretical viewpoint, it is best to use outcome indicators: they measure the impact of care on an individual's state of health. These outcome indicators should be specific to each hospital activity. Apart from the complexity of the task, reflecting the variety of hospital activities, there are also some severe methodological limitations in taking accurate and regular measurements of the outcome of care. In the first place, a final improvement in state of health may not appear until well after the period of hospitalisation: this implies long-term patient monitoring. Unfortunately, the data needed for this kind of monitoring are generally not in the control of the hospitals, and collecting them requires cooperation by the patient and by other care providers. In the second place, if the outcome is measured by the non-occurrence of complications, and if these are rare, it will take numerous observations to be sure that measurement is sufficiently precise. Finally, there may be some confounding factors that must be ascertained before imputing outcomes to a given establishment. For these reasons, and provided there is enough convincing evidence on the link between a process characteristic and an outcome, we may replace an outcome measure by a process quality measure. There are however no properly controlled studies of overall hospital care available, and the relationship between process and outcome is not always fully established. This means that respect for a process standard is not a full guarantee of the quality of outcome.

The third characteristic of hospital activity is the stochastic nature of care. There is no determinist relationship between the provision of treatment and its effect on the patient. No treatment is 100 per cent effective, and no treatment is risk-free. In other words, the failure of treatment may be due to random or other factors beyond the control of the hospital team.

The fourth characteristic of hospital care is that it involves medical practice under real conditions. Contrary to what happens in a well-controlled, random clinical task, where patients are carefully selected, and at the end of which we can measure experimental effectiveness, hospitals must adapt their reference treatments to patients who are not part of a clinical test. As a result of ethical concerns (such as over depriving a patient of a therapy of proven effectiveness) and practical reasons (the cost of constantly conducting random tests), hospitals cannot always fully respect decision-making rules based on data for which there is widespread published proof. In this case, again, adhering to good-practice recommendations in the scientific literature is not a complete guarantee that treatment will be effective.

1.2. *Consequences for the measurement of health care quality*

The first consequence of these characteristics of hospital activity on measuring care quality has already been mentioned: there is no current system that covers all hospital activity, and inter-hospital

© OECD 2002

comparisons are available only for a limited number of pathologies. It can of course be argued that a hospital that does well in several activities is likely to do so in all, but that remains to be demonstrated. There is also debate over the publication of hospital rankings based on combined performance in several dimensions, such as mortality, reputation and other factors (the "league tables"). The aggregation procedure relies necessarily on a relative weighting of various criteria. In general, the choice of these weightings is done empirically, with no scientific basis, by the team publishing the ranking. At best, the relative weights for the different dimensions may reflect the observed preferences for each of these dimensions among the general public. Moreover, aggregation masks performance differentials in each dimension, and can work in favour of hospitals that are merely average on all criteria. In this way, authors recommend the continued publication of non-aggregated results (Eddy, 1998), in order to provide readers with full information.

The second consequence relates to the constraints that statistical method imposes in producing a rigorous demonstration, on one hand, of a hospital's impact on care outcomes (identification of confounding factors) and on the other hand, of significant differences between hospitals. In both cases, these constraints arise, first, from the size of the universe studied, which means that only high-volume activities are selected, or else that data must be gathered over a long period of time in order to achieve a sufficient volume. There is also a further factor of variation – time – during which new technologies or new practices may appear. This means that the publication of comparative results must be significantly hedged, for example by indicating confidence intervals for published indicator values. Uncertainty over the results is the second reason why hospital rankings are criticized: depending on how accurate the measure is, a great number of hospitals may perform equivalently from the statistical viewpoint, and yet may have different rankings. This will make the published results meaningless to a public that is not sufficiently familiar with statistical method, or that does not take the time to examine the results carefully. It can also undermine the ability of insurers to react, if results do not really make it possible to distinguish between various establishments.

The third characteristic relates to the quality of the information systems that are needed to conduct such analysis. Each hospital has to ensure that the data it provides for comparative analysis are accurate and complete. This means controlling the quality of both internal and external data. Data definitions must be strictly the same from one hospital to the next, or the comparison may be distorted. Kritschevsky *et al.* (1999) have shown how different definitions of caesarean section rates can lead to different conclusions in inter-hospital comparisons. Techniques for controlling the impact of confounding factors, for example patient-risk adjustment for mortality rates, can require very detailed clinical data. Such data are often produced for purposes of the care process, but not always systematically (since they are not all required for the medical decision): producing them may therefore require a specific investment by the hospitals being evaluated. In the absence of computerized medical files that are standard from one hospital to the next, there is also a third constraint on the extent of activities covered by a comparative analysis: that of the investment cost in a new computer system for monitoring all hospital activity at a sufficient level of detail. There is debate in the literature over the comparative advantages of specific databases intended to measure care quality for a precise segment of activity (for example, comparing mortality rates for coronary bypass grafts) and databases of the "administrative" kind, *i.e.* consisting of data that are routinely gathered for purposes other than quality evaluation. For example, databases that produce discharge abstracts for measuring the hospital casemix were first used by Medicare, beginning in 1989, to publish casemix-adjusted mortality rates. In the first case, this offers the advantage of data that are specific to a field of pathology, and hence clinically relevant. The results are more convincing, easier to interpret and more acceptable to clinical practitioners. But this comes at the price of an additional investment and a reduced field of analysis. In the second case, in principle, the initial investment cost in the database is low, the field covered is potentially large since it embraces all hospital activity, the data are produced on a regular basis, but the analysis loses its specificity.

2. The main indicators

It should be noted at the outset that this section deals only with indicators that relate to hospitals and that are made available to the general public. Medical practices are also evaluated in many studies that are published in scientific reviews or that are used by hospitals at their own discretion to monitor

© OECD 2002

the quality of the care they provide. We shall adhere to the conventional distinction between outcome indicators and process indicators.

2.1. *Mortality rate as an indicator of quality of care*

There is *a priori* unanimity on this indicator: death is an event that everyone wishes to avoid, and that can be seen as the ultimate sanction of poor-quality care. It has a strong hold on the public imagination, and when premature or avoidable mortality seems to be linked to dysfunctional procedures, it appears particularly unfair. It is also an indicator about which there is no ambiguity in terms of clinical definition, and it is one that hospitals must measure with the greatest precision, if only because the regulations so demand. Since the first publications from Medicare, and from several American states, a comparative analysis of mortality rates has given rise to many methodological studies, and continues to be used by the media in preparing their "league tables" (Ansari *et al.*, 1999; Chalé and Naiditch, 1999; Iezzoni, 1997; Iezzoni *et al.*, 1995; Landon *et al.*, 1996; Leyland and Boddy, 1998; McKee and Hunter, 1995; Thomas and Hofer, 1999).

In fact, mortality data are not that easy to collect. In the first place, many studies show that the records kept by hospitals are not fully reliable, and that they may show significant and unexplained fluctuations over time. Since hospital deaths are a rare phenomenon, missing data can have an important impact on the outcome. In the second place, in-hospital mortality is not a sufficient basis on which to judge the impact of care on a patient 's future prospects. On one hand, the ultimate effect of an intervention may occur well after the first episode of treatment, and the patient may die later at home, or in the course of re-hospitalisation. An acute-care hospital may also have an early-release policy of sending patients to follow-up care facilities, and death that eventually results from the first intervention may occur only in that latter facility. For this reason, it is common to measure not only in-hospital mortality but also mortality within 30 days after discharge (or longer, if relevant), which means that hospitals must monitor their patients after they leave. This requirement can further undermine the accuracy of data if such tracking is not fully reliable.

The remarks above relating to the characteristics of hospital activity also apply with respect to mortality rates. For most pathologies treated in hospital, mortality rates are very low. If we seek to measure them accurately for a given hospital, or to make a useful comparison between hospitals, we must start with a fairly large population. This can be done in two ways: by selecting frequently occurring pathologies, or by selecting data over a period longer than one year. Yet when working with common pathologies, these will often produce a highly favourable prognosis, and hence few deaths.

Moreover, the literature on the measurement of mortality rates stresses the need to take into account confounding factors, and in particular the risk factors that patients present at the time of their admission. Hospitals may differ widely in terms of the profile of the patients they admit. The first adjustment factors to be considered are age and sex. The natural probability of death increases with age, of course, and there is also a difference in mortality at the same age between the two sexes. But other factors must also be taken into account. For any given treatment, patients will present different risk profiles depending on their state of health. The nature of these risks will vary depending on the treatments and the pathologies involved, which means that we must develop specific models and collect specific data for each one.

The models are built on the following principle. An initial medical examination will identify potential risk factors, the impact of which on mortality will then be tested. A multi-centric database is then constructed for hospitals, either on a geographical basis (all hospitals providing treatment in a given state, for example) or on a voluntary basis, or, more rarely, using a predefined stratification. These data can be used to identify significant risk factors and to estimate their relative impact on mortality or on morbidity. This model is then applied to patient characteristics for any given hospital, so as to calculate the expected mortality rates. Next, this expected rate is compared against the rate observed in the hospital, to produce a ratio between the two values. If the observed rate is higher than the expected rate, this may signal a quality problem, while an observed rate lower than the expected rate is

© OECD 2002

a sign of good performance. Generally, it is these observed rate/expected rate ratios that the media use in their published "league tables" for ranking hospitals against each other.

Given the statistical uncertainties surrounding the results, these rankings are unreliable. Observed rates, after all, are derived from a sampling method that will determine the accuracy of the measurement. The coefficients of the statistical models used for calculating the expected rates are themselves estimated, and their accuracy is only relative. It will depend as well on the size of the database and its composition: calibration errors can be suspected if testing the same model on a different basis produces different results, *i.e.* if the risk factors and the coefficients are not the same.

For these reasons, experts have replaced ranking by identifying hospitals that are significantly different from others, the "outliers". This term reflects the notion that the discrepancy between the observed rate and the expected rate for some hospitals is too great to be due solely to chance. An outstanding study of this kind was conducted in New York State hospitals to measure mortality associated with coronary bypass operations (Hannan *et al.*, 1990, 1994, 1995; Naiditch, 1999).

Despite these precautions, the confidence that experts have in the mortality rate (unadjusted and adjusted) for identifying care quality problems is limited (Ansari *et al.*, 1999; Chalé and Naiditch, 1999; Iezzoni, 1997; Iezzoni *et al.*, 1995; Landon *et al.*, 1996; Leyland and Boddy, 1998; McKee and Hunter, 1995; Thomas and Hofer, 1999). There are several reasons for this. In the first place, there are many alternative systems for adjusting risk factors. Iezzoni (1997), Iezzoni *et al.* (1995) and Landon *et al.*, 1996) compared the performance of 12 of these systems for four pathologies or treatments: coronary bypass, myocardial infarction, pneumonia and unoperated cerebral stroke. While the systems' statistical performances are fairly comparable overall, comparing them in pairs produces significant discrepancies, particularly for establishments ranked in the best and worst deciles. In other words, the placement of the "outliers" is unstable from one ranking to the next, even if the rankings are relatively consistent overall. Thomas et Hofer (1999) arrived at identical results, but using a different method. They showed that even if there were a perfect method for adjusting the case mix and the risk factors, systematic errors in the statistical identification of the outliers are such that their positive predictive value is low: in other words, a significant percentage of good-quality hospitals will be found amongst the worst outliers, and *vice versa*. This predictive value increases, nevertheless, with the size of the data samples used for each hospital, but even with samples of 900 patients per hospital the error remains sizeable. Silber et Rosenaum (1997) provided a conceptual critique of the link between mortality rates and quality of care, arguing that not all deaths can be blamed on the quality of care. The general condition of certain patients is such that death is inevitable, and care will merely delay it. On the other hand, the care team must be careful to prevent complications of a condition that could worsen the patient's chances of survival. The rate of complications during hospitalisation is therefore the preferred indicator for measuring quality of care, and it offers two additional advantages. Complications are more frequent than death, and hence the sample size required is smaller. Data are normally gathered systematically from the discharge abstracts, although here the authors call for great caution in using data of this kind.

The mortality rate, then, is a simple and obvious indicator of the quality of care, but it is complicated to calculate with sufficient rigor to be usable and has only weak powers of discrimination between hospitals, once measurement biases are taken into account. Measuring the adjusted mortality rate for one or several pathologies is not sufficient grounds for an overall judgment of the hospital (Rosenthal *et al.*, 1997, Iezzoni *et al.*, 1995). It is not very useful in producing ordinal rankings of the "league tables" type. Finally, it is not clear that it is a good predictive factor for the quality of care, since a high (low) mortality rate is not always synonymous with a poor (good) quality of care. Yet the cumulative and rigorous work that has been done to measure the mortality rate has produced useful lessons: the effort invested in examining patient files so as to ensure the quality of the data and to interpret the results in terms of identifying risk factors has led to progress in reducing mortality rates for certain pathologies, as can be seen from experience with coronary bypasses in New York State (Box 2).

© OECD 2002

Box 2. **The most frequently studied mortality rates**

- Overall case-mix-adjusted mortality
- Mortality following a coronary bypass
- Mortality following a myocardial infarction
- Mortality following severe heart failure
- Mortality following a stroke
- Mortality of patients suffering from pneumonia
- Mortality of patients suffering from an obstructive respiratory illness
- Mortality following intestinal haemorrhage
- Mortality following a total hip replacement
- Mortality of resuscitation patients
- Mortality at birth of underweight infants

2.2. *Intermediate outcome indicators*

This notion covers several phenomena. The result of any portion of the care process is an intermediate outcome. We also speak of intermediate outcome when we do not know how to cure an illness, but can only stabilize certain of its manifestations. In the area of cancer pathologies, a reduction in the size of a tumour is an intermediate outcome. We also measure an intermediate outcome when the final outcome is observed only after a period of time much longer than the stay in hospital. An example is the time it takes for elderly hip-fracture patients to walk again: the final outcome will be observable, not upon discharge from the hospital that performed the acute treatment, but only after a period of rehabilitation. Finally, the frequency of avoidable, undesirable events is also an intermediate outcome: this is the case with nosocomial infection rates, *i.e.* hospital-acquired infections, the complications rate from specific surgical operations, or the importance of iatrogenicity linked to medication prescriptions. Measuring the frequency of avoidable morbidities represents a large part of the work on intermediary outcomes.

The measurement of intermediate outcomes has been the subject of many studies in recent years, under the heading of "outcome research". In the United States, the Joint Commission for Accreditation of Healthcare Organizations publishes a document listing available tools and methods (the ORYX Initiative). Yet these are little used in inter-hospital comparative studies, even though they provide users with more specific information by identifying hospitals' performance for different pathologies. There are several reasons for this.

As in the study of mortality rates, a comparative analysis of intermediate outcomes between hospitals requires some significant methodological precautions, the main ones of which were discussed earlier: sample size, adjustment for patients' characteristics, data quality, need to use specific clinical or biological data surveys. This complicates the use of comparative data. Secondly, the very variety of the hospital workload is an obstacle to generalizing the studies to a large number of establishments.

This has led to an emphasis on performance indicators that are generic and that can be applied to the entire structure. The rate of nosocomial infections has the advantage of being an intermediate outcome indicator that is not specific to a given pathology. It measures the incidence of hospital-acquired infections in comparison to the volume of hospitalisations, and without any direct relation to the originating illness. This rate is complex to measure, since it requires very rigorously documented medical files. Nosocomial infections may be due to staff negligence or failure to respect hygiene regulations, but there is in fact an irreducible risk that infectious agents will be transmitted in any place

© OECD 2002

as open as a hospital. Finally, a nosocomial infection may be due to technical infrastructure shortcomings that are not readily apparent, as in the case of "legionnaires' disease". Hospital staff may see detailed documentation of cases as a kind of self-incrimination, and they may resist keeping records if the episode ends without major harm to the patient. This suggests that the results of incidence surveys should not lead to direct individual sanctions (or rewards), but rather to efforts to improve the functioning of the service. Otherwise, the data may be falsified. This is particularly a problem if the indicator is to be given wide publication, in which case there will be fears of user complaints and lawsuits.

Nosocomial infections are in fact a good example of the way hospitals should account for the quality of their risk management. Effective surveillance of nosocomial infections requires proper observance of very strict hygiene rules, of course, but it also implies a costly mechanism for tracing sources of infection. Yet the enforcement of rules requires a high degree of confidence among the different people involved in hospital care, so that the controls necessary for surveillance are seen not as a threat of sanctions but as a contribution to improving the overall effort and securing the best outcomes for patients. Only under such circumstances will infections be property documented. In other words, a mediocre hospital could well report low rates of nosocomial infections (because few are reported), while an efficient and transparent establishment could find that the data published on it are much less flattering. If infection rates are to be compared among hospitals, then, we must also ensure that each hospital is organized to monitor this risk effectively (Humphreys et Emmerson, 1993; DREES, 2001).

The number of therapeutic accidents resulting from medication prescription errors is an intermediate outcome indicator with characteristics similar to those of the nosocomial infection rate. Personal responsibility is more important here than in the case of infections, however, where it is impossible to control all the infectious agents present in a hospital. For this reason, observation of these accidents is even more difficult. They are regarded as one of the major sources of medical error, and their cost is considerable. Yet to date they have been the object of no published inter-hospital comparisons.

2.3. *Process indicators*

Process indicators are intended to describe and evaluate the process of providing care. They may refer to a practical rule – "measures to prevent thromboembolism should be taken during orthopaedic surgery" – and the indicator will then show the rate of adherence for the rule. In the best of cases, the rule will have been established as a result of experience demonstrating that a given practice is superior. In many cases, there may be no comprehensive experimental evidence, but rather a collection of elements. Finally, sound practice can be dictated by experts. Thus we may use process indicators as a substitute for measuring the intermediate or final outcomes on which they have an impact. In this way we get around the fact that certain undesirable events occur only rarely.

The care process involves a number of phases, and many different players. We need then to identify the stages of the process that have the most significant impact on care outcomes. This analysis can be used to define sentinel events that will signal that a critical event is imminent, one that will pose a high risk to the quality of care. Identifying these signals is important for day-to-day care management, yet given their number it is impossible to make them public in a systematic way. Attempting such effort would be equivalent, for example, to continuously posting the temperature of molten steel in a foundry, and it would have no meaning for users. This is why, once again, research is directed towards finding simple process markers that are generic and easy to collect. Some of these are clinical in nature, and others are not.

The rate of caesarean sections, and the rate of vaginal births after a first caesarean section, are clinical process indicators. We may also consider them as intermediate outcome indicators, where the final outcomes are the state of health of infant and mother. It has in fact been regarded as a characteristic indicator of the care process, since it involves one method of childbirth. The caesarean rate is an indicator of the quality of care, for this reason: a caesarean section is a surgical procedure that requires an anaesthetic, thereby posing a greater risk to mother and child than does natural childbirth. There are

© OECD 2002

medical indications for the caesarean approach, but with the number of these procedures in developed countries growing (admittedly at different rates) experts have come to question the justification in terms of surgical and anaesthetic risk. Studies have shown that, beyond medically justified indications, caesareans may be used as a birth programming technique; that the abuse of artificial labour inducement can also result in the need for a caesarean section; and finally, in certain countries, that caesareans are used as a very indirect form of contraception, based on the belief that a second pregnancy following a caesarean delivery will be dangerous for the mother, because of the weaknesses caused by scarring of the uterus. In all developed countries, reducing the rate of caesarean sections has been accepted as a public health objective. Moreover, the belief that post-caesarean natural childbirth is dangerous is being countered through training and information for professionals and users alike.

We may say, then, that there is a true consensus on this performance indicator. Nevertheless, methodological precautions are needed if we are to draw valid comparisons between institutions. On one hand, maternity ward admission patterns may differ in terms of the maternal risk profile, and hence the probability of resort to caesarean section for valid medical reasons. Comparisons must therefore be adjusted for these risk factors (Aron *et al.*, 1998). On the other hand, Kritschesky *et al.* (1999) showed that caesarean rate calculation was dependent on conventions relating to the information-keeping systems of each hospital, and that no absolute inter-hospital comparison was possible.

The most frequently used nonclinical process indicators are length of hospitalisation, rates of unplanned readmission, and volumes of activity by type of intervention. Certain magazines that publish "league tables" make the assumption that a short stay in hospital is a sign of good-quality care. On one hand, this is what patients want. On the other hand, a short stay reduces a patient's risk of exposure to prescription errors and to nosocomial infections. This pressure towards shortening the length of hospitalisation also has an economic rationale, in terms of reducing the number of hospital beds. But care must be used in interpreting it. A comparison of length of stay requires adjustment for different factors: the case mix, the severity of cases, and sometimes the socioeconomic status of the patients. On the other hand, we must be careful not to invert the causal relationship. It is true that new treatments have appeared that reduce the need for long hospitalisation, such as endoscopic exploratory and surgical techniques, while at the same time enhancing the quality of care. In this case, it is the improved techniques themselves that have reduced the length of hospitalisation and increased the quality of care. Yet reducing the length of hospitalisation is a parameter that can be manipulated by the hospital. If the hospital has surplus beds downstream in its convalescence or rehabilitation wards, early discharge from acute hospitalisation may lead in the end to a longer convalescence and a less satisfactory quality of care. For this reason, the length-of-hospitalisation measurement must be accompanied by a measurement of the number of premature discharges, and the number of readmissions following on the heels of such discharges (generally within 30 days). In fact, a high number of early readmissions may indicate inadequate patient rehabilitation. This measure is not in itself completely satisfactory: some hospitals may refer their patients to another hospital if they suffer post-operative problems.

Using the workload by type of intervention is based on a simple idea that originated in industrial economics: cumulative experience in executing a task improves the performance of the operator. Several studies have attempted to discover this "experience effect" in the health system, particularly for surgery or diagnostic tests (Hillner *et al.*, 2000; Nathens *et al.*, 2001; Phillips *et al.*, 1995; Ritchie *et al.*, 1999). Since the intent is to show the effect of experience not only on the time spent in surgery, or the cost of the procedure (primary productivity), but also on the quality of the outcome, measured for example by the number of peri- and post-operative complications or deaths attributable to surgery, it is important in these studies to observe the same methodological precautions as in other performance measurement research. In particular, patients cared for by different surgeons must have identical risk factors, and the surgical technique employed must be the same. Some studies look at cumulative experience over time, and this raises the problem of the appearance of new techniques. In the end, the relevant level of analysis is not solely the intervention of the surgeon, but the overall volume of a team's activity. A team that does few operations of any one type but that has a high overall level of activity will have more experience than a team that does few operations of any kind because its overall workload is low. Finally, the analysis must apply to the entire operating team and not only the surgeon.

© OECD 2002

This experience effect has been applied (with due regard for the above factors) to certain surgical procedures, such as total hip replacement, peritoneoscopic surgery and open-heart surgery, to mention only a few. The volume of activity by type of operation is often used as a process criterion in inter-hospital comparisons, and particularly in the preparation of rankings. Yet these rankings should not be interpreted too literally. In the first place, studies show a significant statistical link between volume and outcome, but this link is not enough to move directly to a ranking. Suppose, for example, that 30 per cent of inter-hospital performance discrepancies can be explained by the volume effect. The fact remains that 70 per cent of the differential must be explained by other factors. To jump directly to a ranking will give rise both to positive errors (where hospitals may be highly ranked on the basis of their high volume, without deserving it) and negative errors (where hospitals are poorly ranked, and improperly so, because of low volumes). Consider for example a surgical team that performs a high volume of operations, but at the price of overworked support personnel and long days in the operating room for the surgeon. Studies have shown that when surgical teams approach the saturation point there is a high risk to the quality of their work, because of fatigue and inattentiveness. Another reason for avoiding ordinal rankings is that volume effects are not continuous: a hospital that does 152 operations a year is not necessarily performing better than one that does only 151. It is important therefore to establish workload thresholds.

In terms of the information that should be given to users, an indication of volume is not sufficient. We must also indicate whether the operating team follows the rules in terms of the pace of work and recovery time. We must also know if other, more specific process criteria are being respected: for example, the systematic use of prophylactic precautions during intestinal surgery to avoid post-operative infections. These process indicators are generally studied in the course of accreditation. This opens the question of the publicity given to these reports, which contain the most extensive available information on a hospital 's working methods.

2.4. *Lessons*

Several lessons can be drawn from the experience described above in terms of a comparative evaluation of hospitals. In the first place, all authors insist on the issue of the quality of information systems and the degree of standardization required to ensure that studies are conducted rigorously.

Investment in the establishment and maintenance of databases is thus a cost that must not be underestimated. Nor should we underestimate the skills that may be required to manage them and to analyse the data. We have already pointed to the methodological precautions that must be taken in order to reflect confounding factors in inter-hospital comparisons. Consequently, any organisation, public or private, that seeks to develop an analysis of this kind will have to recruit high-level statisticians and enlist the cooperation of expert clinicians in interpreting the results.

The last point concerns the publication of hospital rankings in the form of “league tables”. The media are fond of presenting results in this way, because it is simple and attention-grabbing and good for circulation. But there are sound reasons for thinking that these rankings are spurious, in the sense that they do not really reflect inter-hospital differences in the quality of care. We have discussed these reasons: the results are tainted with uncertainty and the indicators are not direct and unequivocal measurements of quality. A further reason is that these publications attempt in the end to impose a weighting on the different dimensions of quality, for example by giving more weight to mortality than to length of hospitalisation. Yet this weighting function has no justification in science, in politics (the results of a vote) or in economics (marketing surveys that might determine consumer preferences). The final ranking is thus sensitive to the choice of weightings.

Such information is supposed to help users to become enlightened consumers, so they can choose the hospital where they will go for treatment. If the information is erroneous, then the media are providing poor guidance for these decisions and could, in theory, be responsible for severe patient disappointment. They might even find themselves liable to lawsuit for providing false information occasioning personal injury. Low-ranked hospitals might also consider themselves libelled. At the very least, it would be wise to publish, not ordinal rankings, but comparisons between groups of hospitals with statistically comparable levels of performance.

© OECD 2002

3. The impact of comparative evaluations

We shall analyse this impact from three viewpoints, that of the providers of health care, those who use them, and those who pay for them. There has in fact been little effort to evaluate the impact of these inter-hospital comparisons.

3.1. *Impact on care providers*

The publishers of these data expect that the impact will be to improve the performance of each care provider. This improvement will itself depend on the structure of incentives contained in the original program. In a conventional competition model, the provision of information on product quality does not necessarily produce an incentive to low-quality producers to improve their performance. In fact, they may conclude that some consumers are price-sensitive and willing to settle for lower quality. These producers will be able to find their own market niche. At the very most, there may be regulations establishing a minimum quality threshold below which no one may operate. Why does any producer seek to improve his quality? He will do so if he thinks he can earn a return on his investment, through a larger market share and higher profits.

In the health care sector, we must anticipate, then, that the publication of comparative data on quality of care will have a quality-improvement impact only on those providers seeking to expand their market share and hence their revenues or profits.

The health-care industry has a number of specific features. One of these is the high price/quality elasticity of users, *i.e*. users accord great value to quality and are ready to pay much more in order to guarantee that quality.[4] In other words, publishing comparative results could provoke significant shifts of clientele between establishments (or, in the case of systems where there is heavy litigation pressure, such as in the United States, the fear of malpractice suits could induce the poorest-ranking hospitals to improve their performance). If hospitals feel that this threat is high, they will have an incentive to invest in quality. Yet the perception of this risk is tempered, in the case of hospital care, by the issue of geographic accessibility (how far people are willing to travel in order to obtain a higher quality of care) and the degree of competition within the local catchment area (a hospital with a local monopoly will have no incentive to change). Finally, for any given hospital, we may assume that the decision to take action is a function of perceived cost in comparison to expected returns, and this is also a function of the gap separating the hospital from the average of the hospitals with which it is being compared. If this gap is narrow, the incentive to take action is likely to be weak, and the hospital can ascribe the discrepancy to statistical error. For hospitals ranked above the mean, we may also assume that they will have little incentive to try to catch up with the leaders if the perceived cost of such an effort is high.

A second specific feature of the health-care industry is the involvement of a third-party purchaser, the insurer (public or private). The price/quality relationship is negotiated between the insurer and the care provider. In the case of private insurers in a competitive market there are two possibilities. The insurers may deem the price/quality relationship to be satisfactory (for example quality and price are both low), in which case there will be no incentive to change, in the absence of evidence that the insured clients are unhappy with the situation. If an insurer feels that that the price/quality relationship is inadequate, it may threaten to withdraw its accreditation of the hospital, or reduce the price it will pay. This threat will be all the more credible if the insurer is a dominant source of financing for the hospital, or if all insurers adopt the same strategy. In this case, the hospital will be forced to take steps to improve the quality of the care it offers.

In the case of a public insurer, we may assume that it is in a monopoly situation. In this case, it is politically difficult not to take action when quality discrepancies between the various hospitals under its control become too great. Here we may identify four possible situations:

Price/Quality	Low	High
Low	1	2
High	3	4

© OECD 2002

The public insurer may find that there are hospitals 1) of low quality and low unit cost, 2) of high quality and low unit cost, 3) of low quality and high unit cost, and 4) of high quality and high unit cost. In 1), it should in principle condition its financing of the hospital on steps to improve quality. Those establishments that fall under 2) will constitute a reference group. Hospitals in 3) should ultimately face financial penalties, with demands that they improve their quality. Those in 4) should be asked to reduce their unit costs while maintaining their current quality levels. There is no incentive, then, for a hospital to make an effort unless the purchaser has the technical means to perform a unit cost/quality analysis, and the political clout to enforce it. In fact, this means a redeployment of resources among hospitals, which may mobilize their local political resources to influence the decisions taken.

In short, for a hospital with a mediocre record, the incentive to act arises from *a*) the fear of acquiring a negative reputation that will over time result in fewer patients; *b*) the quality gap by which it falls short of the average hospital; *c*) this fear may be tempered in the short term by the hospital's relative positioning in its local competitive environment; *d*) it also depends on the pressure exerted by the insurers; *e*) this depends on the insurers' perception of the price-quality relationship.

This means, in effect, that the reaction to published results is likely to depend on the particular circumstances of each hospital, and that it is difficult to predict any general tendency. Yet this reasoning holds only at a given point in time. If the results are published repeatedly, showing that certain hospitals are making no effort, then they will sooner or later find themselves obliged to take corrective action. In terms of measuring the impact of comparative evaluations, this means that the effort must be maintained over the long term. In the case of a public purchaser, however, it must ultimately accept responsibility for correcting discrepancies, especially in the case of hospitals where quality is low and unit costs are low.

Does the literature back up these predictions? The evidence is too fragmentary to permit definitive conclusions, but we may discern some indications. The experiments cited as having had their impact on the behaviour of care providers evaluated are based primarily on voluntary participation by hospitals, which in fact pay for a portion of the studies. This means that those hospitals consider *a priori* that gaining recognition of their relative performance is sufficient benefit to offset the risk of a negative image, assuming that consumers and insurers will give them credit for having taken the risk of being shown to be underperforming. They are thus in a position where they must make every effort possible: if their results are mediocre, they will have to take action so as not to lose the confidence of their respective clientele.

One of the first large-scale experiments conducted in the United States involved the regular publication of casemix-adjusted mortality rates by the Pennsylvania State Department of Health. This experiment went quite far, and involved the publication of physician-specific data. Launched in 1989, it was suspended in 1993 after a series of disputes over the quality of the adjustments made and the appropriateness of using mortality as a performance indicator.

The most famous example in many respects is the coronary bypass mortality study in New York State (Hannan *et al.*, 1990, 1994, 1995; Chassin *et al.*, 1996; Naiditch, 1999). This was an exemplary study on several scores. It involved all the New York State hospitals performing coronary bypasses. It was financed by the state Department of Health, which committed itself to long-term funding. The study lasted from 1990 to 1997. It gathered specific clinical data involving long-term patient tracking, and data quality was monitored regularly. Results could thus be properly adjusted for risks. The teams met frequently to analyse the results. Results were presented, not as rankings, but as a three-tiered grouping of hospitals: those with better-than-average scores, those with scores around the average, and those below average. Initially, to encourage surgeons to sign on to the project, it was agreed that the results would not be published until the entire study was completed. Partway through the study, however, a journalist, claiming freedom of information, challenged this understanding in court. Following the first publication of results, mortality declined by 41 per cent, while it went down by only 18 per cent in the rest of the country, and patient severity rose. Part of this improvement was due to the fact that the less successful teams, in general those with the lowest workload, abandoned the market,

© OECD 2002

but the systematic study of the most successful working methods also helped other teams to progress, and to understand the factors for success with the treatment.

Another large-scale experiment that gave rise to a number of published studies was conducted in Ohio, under the name Cleveland Health Quality Choice (CHQC) (Rosenthal *et al.*, 1998). This project dates from 1988 and represents a joint effort by several groups of stakeholders: the hospital association, the physicians' association, employers and the chamber of commerce. It assembles data from 30 hospitals, covering the following areas: mortality associated with eight pathologies, length of hospitalisation for 13 pathologies, mortality and length of hospitalisation for resuscitation patients, rate of caesarean sections and rate of vaginal births after caesarean section, indicators of hospital iatrogenicity, and a survey of patient satisfaction. The published studies on the impact of this program show a decline in the caesarean section rate, an increase in the rate of vaginal births after caesarean section (Caron and Neuhauser, 1994), a reduction in post-resuscitation mortality rates together with a reduction in the length of stay, and this despite a heavier patient load (Sirio *et al.*,1994). Rosenthal *et al.* (1998) cite four case studies of hospitals participating in the experiment that took steps to improve their results in three areas: length of hospitalisation for coronary bypass, pneumonia-associated mortality, and the caesarean section rate and the rate of vaginal births after caesarean section. An external evaluation of the Cleveland experiment was published by Rosenthal *et al.* (1997), who examined the mortality rates for the eight pathologies selected at the outset of the experiment and performed a statistical analysis comparing these rates with those before and after the CHQC. They concluded that there was a significant reduction in risk-adjusted mortality for the eight pathologies, totalling 13 per cent over two years, with a 20 per cent improvement for heart failure, intestinal haemorrhage, colectomy and coronary bypass. Given the nature of these data, the analysis does not permit a solid inference that this trend is attributable to the project. A very similar experiment was undertaken in California, known as the California Hospital Outcomes Project (Romano *et al.*, 1995). This project was limited to three problems: mortality after myocardial infarction and after vertebral disk surgery, and the caesarean section rate. Finally, Epstein and Kurtzig reported in 1994 that 38 American states had established databases to assess hospital performance (Epstein and Kurtzig, 1994).

The Cleveland experiment is instructive on several points. Initially based on voluntary participation through a steering group that included all partners in collective decision-making, it experienced a number of tensions over the release and use of results. One partner used published results to renegotiate its contracts with the hospitals, but without notifying the other members, and the hospitals threatened to withdraw from the project unless there was proper consultation before any of the data were used for taking decisions. Secondly, the program was limited to producing comparative information, but it was left to the initiative of individual hospitals to take improvement measures. The project group never provided hospitals with any assistance in these actions. Thirdly, the hospitals shouldered 75 per cent of the overall cost of the project, amounting in 1998 to some $3 million for the 30 hospitals, or about $100 000 each.

In the United Kingdom, Bucknall *et al.* (2000) showed that introducing a simple system for auditing the care accorded to asthma patients resulted in closer observance of good-practice recommendations.

In conclusion, the published data suggest that the establishment of comparative databases on hospital performance, if they are constructed with rigorous methodological principles and based on cooperation with the participating hospitals, can provide highly useful knowledge for improving performance. Publishing the results is an essential step, but the potential threat that this represents for participating hospitals must be tempered by predetermined rules governing use of the resulting information, as well as the publication format. The authors, here again, stress the difficulty of producing high-quality data. We can identify efforts that hospitals have made to improve their performance in the wake of published evaluations, but these are not systematic to all institutions. It would seem, in the end, that there is an overall effect, but it is difficult to attribute this to the comparative analysis alone. It is likely that these improvements are also the result of a more demanding external environment for hospitals.

© OECD 2002

3.2. *Impact on users*

The publication of hospital performance data will have an impact if users use this information in selecting their care provider. From an economic theory viewpoint, such publicity is justified because it reduces the asymmetry of information between service providers and beneficiaries. From a public policy viewpoint, this publicity is desirable as a means of bringing the needed transparency to public services. These principles are hard to contest. In practice, patients suffering from certain chronic diseases (such as renal failure or HIV/AIDS) have demonstrated that they are able to acquire some real practical knowledge about their disease and the treatments they need. They are in a position to truly express an autonomous demand for health services. But these expert patients are not the rule, and in most cases demands for health care are not planned carefully ahead of time. The question that arises then is this: what is the best way to publicize this information so that it will be truly useful to beneficiaries of the services, and will not lead to mistaken judgments? From this viewpoint, debate over the publication of "league tables" takes on a new perspective. These rankings offer a highly simplified account of relative hospital performance. As we have seen, ranking hospitals in terms of mortality rates, for example, has no scientific value and transmits spurious information. "Sound" information is that which groups hospitals with similar performances (within a single confidence interval), yet even this kind of grouping is not error-free. In other words, selecting a hospital from a list simply because it is in first place is no guarantee that it will give a better quality of care than the second or third-placed hospital. It may even present a risk of poor-quality care. This "true" information may be too complicated to be acceptable, however, and may even undermine a lay reader's confidence in the researchers' work.

This information should be useful for the following kinds of decisions:

- In the case of a sick person, what is the treatment best suited to his condition?
- Which service providers offer these treatments, and at what level of quality?

We have discovered little research on this issue. We do not know the initial public impact of publishing comparative data: how many people read them, with what degree of attention, and what credibility they may accord the data provided. Voelker (1995) notes that data from the accreditation surveys conducted by the JCAHO are available to the press and to the general public, but they are seldom consulted. Even when this information is accessed, we do not know what role it may eventually play in individual decision-making. For example, we have little evidence on the impact that a published hospital ranking may have on admissions in a given hospital. On the other hand, Hibbard *et al.* (1996) conducted studies in order to 1) evaluate people's understanding of quality indicators and 2) describe the most effective means of communicating results. Their work attempted to assess the impact of comparative information in terms of facilitating the choice of a health insurer. Their findings can be summarized as follows.

In the first place, these authors point out that published studies are not the only source of comparative information available to beneficiaries. Competing or supplementary sources include past experience with a hospital, word-of-mouth, the personal physician's opinion, surveys published by patients' or consumers' associations, and other sources of information offered by the media. Measuring the impact of published data requires complex research tools for observing the decision-making behaviour of individuals in a "realistic" situation, either through experimentation or through interviews. No such studies have yet been performed.

Jewett and Hibbard (1996) highlighted the complexity of information on the quality of hospital care. Using the focus group method, where a limited number of individuals are asked to pass judgment on assertions or to respond to questions, and where group dynamics encourage people to speak out more freely and to be more explicit in their responses than they would in a questionnaire-based survey, they found that a significant percentage of people had a poor grasp of what the indicators mean, and that they either misinterpreted them or did not understand them at all. This was particularly true for indicators relating to undesirable events. Under these conditions, we must ask how we can present data so that people will accept them as reliable, *i.e.* how to indicate their degree of accuracy. Hibbard *et al.*

© OECD 2002

(1997) suggest the need for efforts to explain the data and place them in context. People at whom the information is targeted should be told why a certain indicator has been selected, how it actually measures the quality of care, and what questions it can answer. This exercise can become very difficult, however, especially if the indicators selected are published with the statistical notes demanded by sound methodology. Finally, the authors note that the relative sophistication of published information makes it inaccessible to people with a low level of education.

Aggregating the different dimensions of performance poses another problem. In the studies published by the mass media, this issue is resolved through an arbitrary "in-house" weighting of these different dimensions. Yet without such aggregation, the reader himself will have to impose his own trade-offs between the various dimensions of quality. According to Hibbard *et al.* (1997), empirical studies on decision-making show that individuals find it very difficult to make trade-offs between conflicting dimensions. In this case, they will try to construct an analytical framework that will favour one dimension over the other, at the risk of underestimating or ignoring significant pieces of information.

For all these reasons, and in an effort to improve the decisional utility of the information provided to patients, Hibbard *et al.* (1997) and Sisk (1998) propose assigning an important interpretive role to mediators, who should ideally have three types of skills. First, an understanding of the questions that patients ask before deciding on hospitalisation for a given pathology: these questions are likely to concern the procedures involved in the treatment as much as the risks associated with it. Secondly, an understanding of the way in which hospitals actually deal with these problems: this will require further and more refined studies of process, hospital by hospital. Finally, an understanding of the use and limitations of available outcome indicators that make it possible to explain a situation and place it in context. Patients' associations might play this role, as well as people's personal physicians, provided they have access to all this supplementary information. Publishing comparative data would therefore mean not only printing tables of results in the press, but also giving people access to a network of accredited intermediaries for advice.

Finally, there is much stress on the credibility of the information produced. This credibility has three dimensions: the independence of the producers of comparative information, the reliability of the data (which is linked to the degree of standardization in calculating the indicators) and their accuracy (the quality of the information systems). In an effort to certify this credibility, given the profusion of initiatives for publishing comparative results among care providers and health insurance plans in the United States, an evaluation and accreditation body, the Consumer Assessment of Health Plans Study, was created in 1996.

3.3. *Impact on the purchasers*

This point was already addressed in part during our discussion of care providers. In principle, awareness of hospitals' relative performance in terms of care quality should have an impact on their financing, as a function of the cost-quality relationship sought by the purchaser. We can imagine three decisions that might be related to the relative performance of the care provider: the decision whether or not to send insured patients to this provider, the decision to sign a selective contract with a provider (where the insurer selects a set of treatments from among the full range offered), and finally the negotiation of the price-quality relationship. In this last case, a provider offering higher-than-average quality could also demand a higher price. On the other hand, the purchaser might refuse the price demanded by a provider whose level of care is not up to the purchaser's expectations. Such a trade-off requires the ability to make simultaneous price-quality comparisons among care providers.

Generally speaking, the first decision is taken on the basis of an accreditation process entrusted to an independent agency. Accreditation is in principle the cornerstone of systems for measuring the quality of care. Detailed information on a given hospital will be made available to potential purchasers and, to a partial extent, to the public. Information of this kind will attract less public attention than the league tables, no doubt because the reports containing it are so lengthy and so technical. The care provider's accreditation level is thus an essential parameter of its relationship with a provider. But in a country where accreditation has been in force for a long time (as in the United States), nearly all

© OECD 2002

hospitals are accredited at a satisfactory level, and it is no longer possible to discriminate among them on that basis. There are many examples of selective contracting between purchasers and hospitals: if the level of performance for a segment of activity is not satisfactory, the purchaser will refuse to refer patients needing such treatment to that hospital. In health care systems where there is a strong public sector presence, this is done through selective accreditation procedures: only those hospitals that meet a certain number of criteria (generally process criteria) will be allowed to practice that segment of activity. The segments most often cited are heart surgery, cardiology institutes, organ transplant centres, and maternity hospitals. Outside the United States, these accreditation procedures are based primarily on process criteria, and not on outcome indicators.

Cost remains a dominant criterion of the relationship between purchaser and provider, particularly in competitive health insurance systems. In the United States, where such competition is sharp, the quality of the health plans offered to insurers has become an important business consideration with insurees and their employers, who generally pay for these programs. In systems where the public sector plays a major role, the quality and safety of health care have become a significant political issue. In other words, it would seem inevitable that negotiations between purchasers and providers are going to become an increasingly important part of this dimension of the qualitative performance of health care. The studies published on this topic are intended primarily to evaluate whether pressure on costs leads to deterioration in the quality of care provided to the insured. On the other hand, since these negotiations are conducted in private, there are no published studies for determining how economic agents make their trade-offs between quality and cost.

We may suggest the following hypotheses. First, we may suppose that the unit cost of service rises with its quality: this is the dominant assumption, that quality is expensive. In a system where purchasers are in competition, and where service providers are also in competition, the purchasers have only partial information available on the relative quality of the providers. They have the basic guarantee offered by accreditation. They may also resort to selective contracting. But the published indicators cover only a portion of providers' activities. The purchaser must therefore make the assumption that if a hospital is doing well in one part of its activity, there is a good probability that it will do so in all activities. On the other hand, the indicators make it difficult to discriminate between providers: the league tables are misleading, by any rigorous standard, and can at best classify providers into three or maybe five groups: high-quality, average/acceptable, low-quality. The economic calculation whereby the purchaser will determine the trade-offs between a marginal unit of additional quality and the unit cost of the service is not easy to perform without a specific examination of each provider. Moreover, it can take two or three years of data collection to confirm an initial ranking. Finally, purchasers are subject to geographic constraints: in any given zone, there are only a finite number of providers. It is generally impossible to select a provider from beyond the residential zone of the insured. Under these conditions, purchasers are likely to avoid the lower-ranked hospitals as far as possible, and to make competition work for them (in terms of costs and prices) among those ranked in the middle. On the other hand, they are not likely to have much bargaining strength with the small number of high-quality establishments. There will always be a degree of heterogeneity, therefore, in the quality of care.

In a system of universal health insurance with a single purchaser, that purchaser will face the same difficulties in making fine distinctions between providers on the basis of their performance and their price-quality ratio. Yet the purchaser has to operate under the constraint of treating all providers equally in terms of their quality of care. Assuming that cost increases with quality, then it is the purchaser's responsibility to make more resources available to the lower-quality providers so that they can catch up with the others. If the results are published, the purchaser may be subjected to heavy pressure from voters (*i.e.* the insured) to upgrade the quality of service and hence to provide more resources across the board.

In both cases, average quality will increase along with costs. In the first case, the poorer providers will be forced out of the market if they cannot shape up, while in the second case the public authorities will guarantee to all providers the means to achieve the best level of performance. We may have some

© OECD 2002

doubts about this scenario, however. In fact, the notion that "quality is expensive" may be true once all providers have reached the same level of performance, but that is far from being the case today. In other words, it is quite possible for providers to enhance their performance while reducing their unit costs. For example, many studies have shown that there is a positive correlation between workload and clinical outcomes. In the hospital business, where fixed costs are high, unit costs will decline as a function of volume. Sending patients to high-volume hospitals, then, can serve both to improve clinical outcomes and to reduce costs. The purchaser can make a quality-cost trade-off only if it is in a position to know the relationship between those two factors, activity by activity. We have already pointed out that only a small part of hospital activity is covered at this time by performance indicators, which means that we are surely far from having exhausted the potential for improving outcomes without increasing costs. There is a further difficulty: the people insured by a given purchaser will not necessarily have the same profile as the average for patients of a given hospital. Since performance indicators cover all of these patients, there may be a performance estimation bias in one direction or the other, depending on whether the purchaser's clients present a greater or lesser risk.

We are faced, then, with a sharp asymmetry of information between insurers and providers, to the detriment of the former, in terms of the quality of care, despite the development of comparative indicators. The purchaser must find incentives that will induce the care provider to make visible efforts to improve its results and to share any productivity gains with the purchaser. For these efforts to be monitored, the provider will have to develop an information system that goes beyond producing the indicators most widely used today. Consequently, transaction costs between the two players will have to rise.

We can now see why studies that attempt to measure the impact of competition on the quality of care tend to produce contradictory results, quite apart from the methodological difficulties of measuring this competition (McLaughlin and Ginsburg, 1998). On one hand, it is by no means certain that the purchaser can really observe the provider's effort to improve its performance. In this case, costs may decline and performance may deteriorate without the purchaser's being aware, or the costs of quality may be shifted to another purchaser or another segment of activity where performance is not measured; or there may be cost-reduction gains resulting from improved performance that are not shared with the purchaser. On the other hand, the intensity of competition may vary depending on the segment of activity of each provider.

3.4. *Summary*

The lessons available about the impact of publicizing hospital performance indicators come mainly from United States, where publicity of this kind is a feature of the competition among purchasers to win insurance clients, and among providers to win purchasers. US experiments have been based on the principle of the "informed consumer", to the point where one initiative, the Consumer Assessment of Health Plans Study, guarantees the quality of information provided by the various open health insurance plans. The weight of consumerism in the United States makes itself felt as well, in the role played by patients' or consumers' associations in publishing comparative health-care studies. The climate is not the same in England, where "league tables" have existed since 1993, but where publication of mortality data began only in 2000. Nor is it the same in France, where publication of comparative performance data began in 1999, and remains the preserve of the mass media. Public purchasers in these two countries have not really introduced any mechanisms for care providers to compete on the basis of their clinical performance. The impacts observed in the United States are thus greater than we should expect to find in a country where the public purchaser is in a monopsonistic position.

We find, then, that the publicity given to comparative studies has had an impact on a portion of the providers concerned, but not on all of them. Some hospitals have undertaken corrective actions and have succeeded in improving their performance significantly in particular treatments. These actions have been most systematic when the study was clearly targeted at improving performance in a particular treatment. In other experiments, such as that in Cleveland, the initial objective was rather to

© OECD 2002

make comparative data available to all players as a basis for making enlightened choices, but the initiative to take corrective action was left to each provider. That initiative is no doubt itself dependent on the provider's competitive position, its negotiating strength with purchasers, and the political clout exerted by local consumers' groups. Despite this limitation, these experiments provide important and useful knowledge to all players on the critical factors in clinical performance. Finally, the impact on patients is ambiguous. The published articles show that information of the clinical type is complex and often misunderstood, and the uninitiated are likely to prefer patient satisfaction indicators to clinical indicators. The publicity given in the press to the ratings awards apparently have a short-term impact, but not a lasting one. There are no studies to reveal how the general public has used this information in choosing a treatment or a hospital. This does not mean that such efforts should be abandoned, but rather that more effective and better-targeted means of communication are required.

4. Recommendations

Publicizing comparative results between hospitals, health services and insurance plans would appear to be an irreversible trend, for several reasons. Rationalizing resource use in health systems implies reducing informational asymmetry between the players, to the benefit of patients and the public or private insurers that speak for them. This publicity is equally justified whether in an American-style competitive system or in public systems that are accountable to their citizens for the quality of tax-funded service. The irreversible nature of the trend is reinforced by the growing ease of access to general information on diseases and their treatments, even if such access is still highly unequal. Yet if publicity is to be of real benefit to the community, it must meet requirements in terms of 1) information systems, 2) partnership with providers, 3) communication with the beneficiaries of health services, and 4) formalizing the price-quality relationship with purchasers.

4.1. *Information systems*

All the authors stress that a great deal of work will be needed to ensure *a*) a standardized definition of the data to be collected, *b*) the regular collection of data, *c*) the accuracy of information, and *d*) the quality of statistical manipulation. Such work must be accepted as a responsibility by each provider, of course, but also by the organisation seeking to constitute a comparative database. At the outset, the experiment will require a long training and breaking-in period. During the course of the experiment there will have to be regular audits, with frequent resort to medical files and discussion with the care team to ensure that the information is valid. Each new participant will add to the value of the database, but there will be few benefits of scale when this happens, since while the organisation managing the data may have acquired a degree of expertise, any new participant will still have much to learn. The slightest slippage in terms of quality will throw into question the value of the conclusions and credibility of the database itself.

There is debate over the best way to constitute these databases. In order to reduce the costs of constituting and managing them, some experts recommend making maximum use of existing databases, either administrative or regulatory. The best-known tool is the discharge abstract, which serves among other things to identify a hospital's output in homogeneous patient groupings (the Diagnosis-Related Groups or any other hospitalisation classification system). Yet these data have a drawback in that they are produced for rate-setting purposes, and their coding thus follows precise criteria that may fall far short of a rigorous description of the care process. Moreover, the medical data available are not always sufficiently detailed to measure the outcome of the care process. This is why other experts recommend constituting specific databases, drawing upon detailed clinical data collected during the patient 's period of hospitalisation. In this case, data production is more expensive and data quality is more difficult to maintain. In fact, the health teams must routinely gather data that they will use only briefly in the care process, but that they have no need to memorize. The utility of such an effort is apparent only after the analysis has been done, long after the data are gathered. It is difficult to decide between these two approaches. We do not really know their consequences in terms of costs, compared to the expected benefits of more specific information in the second case. For countries with a shorter history than the

© OECD 2002

United States has in gathering discharge abstracts, it would be a mistake to think that they can produce prompt results with existing data: American-style discharge abstracts already contain more detailed information then those used in European countries. Some authors place great store in the spreading use of computerized medical files. Yet here again, we must not underestimate the difficulty of developing a common structure of medical files for any country, or a standard definition of all the items involved.

4.2. *Partnership with the medical profession*

These databases serve to compare the performance of medical teams. Physicians are of course ready to participate in efforts to improve their results, but they will be suspicious about the penalties, whether economic or legal, that may be entailed. They are particularly sensitive to the image that a hospital may acquire after publication of negative results in the media. Yet they are a critical and indispensable resource for conducting any such studies. It is they who produce the basic data on the system. They are thus in a position to undermine databases by simply relaxing their vigilance over the quality of information produced. Indeed, such behaviour may be justified if the published data are going to be used to bring malpractice suits against practitioners (Liang, 2000). They are also the principal players in interpreting the results and in translating these into their practice. Finally, their cooperation is likely to be required over the medium term in order to produce significant results. The rules of the game in terms of communicating results, and the decisions that flow from them, should therefore be negotiated in advance and respected for as long as the database is in use. In the Cleveland experiment, for example, cooperation nearly collapsed when one of the partners, the purchaser, reversed its initial commitment and decided to make use of the study results for selecting between hospitals, without submitting the issue to the whole group in advance. It was not the decision itself that was criticized, but the failure to respect the group's written rules. This warning should be understood as a plea to accord medical personnel a degree of immunity: it reflects the clear finding that, even if it is easier today to monitor their behaviour and hold them accountable, they still have considerable discretionary power to justify their decisions *a posteriori*. Informational asymmetry may be reduced, but it does not disappear.

4.3. *Communication with patients*

We know little about the impact of rankings published in the press (or by other means) comparing clinical outcomes for hospitals. It is time to examine this question in greater depth. The most plausible hypothesis is that the attention effect will be short-lived, and will fade over time. This is logical enough: the information produced will be useful mainly to someone who is sick or undergoing treatment, if the ranking provides specific information on that disease and its treatment. Yet this is clearly not the most important point. The fact is that information on clinical outcomes is very complex, particularly if it is given with the kind of methodological reservations favoured by statisticians. It may be misunderstood or its meaning may be completely inverted. It often does not meet the real expectations of the general public, who tend to be more interested in indicators that look at the interpersonal relationship dimension in the caring process and patient satisfaction. People with less education might find clinical outcomes data even more incomprehensible. And these data do not cover all pathologies.

What is needed, then, is to find efficient ways of communicating this information. This can be done by improving the form of the published documentation, which should include several levels of communication and explanation ("layered information") so that everyone can delve further into topics of interest. The information also needs to be contextualised, *i.e.* the cognitive space in which it has meaning should be explained. It is not enough to report outcomes: the treatment process must also be explained. The use of mediators – patients' associations, for example – can be effective in responding to people's questions. Finally, the means of communication must be adapted to the target audience. This in effect might mean launching a process of acculturating the general public to information on the clinical outcomes of hospitals.

© OECD 2002

We took a critical look at "league tables" earlier in this paper. Rankings of this kind reflect only very roughly the real differences in performance between institutions, and many institutions could well be entitled to demand amends for attacks on their reputation. It is tempting to reproach the media for sensationalism in publishing data in the form of unvalidated rankings. Yet since we do not really know the impact of such publications on the reputation and attractiveness of hospitals, it is difficult to measure their negative effects. There is evidently an effect on hospitals that feel unfairly treated in the rankings: exercising their right of response will not command the same audience as the initial publication. Three solutions suggest themselves for insuring that publications are more rigorous. First, a competing publication might seek to produce a counter-assessment relying on a more rigorous approach, or demonstrating the progress that may have been made since the previous ranking. Second, dialogue could be encouraged between experts and journalists, leading gradually to the publication of more rigorous data with due regard to statistical limitations (Chassin *et al.*, 1996; Rosenthal *et al.*, 1998). Third, the organisations producing the studies could arrange to disseminate them themselves, with the help of communication experts.

4.4. Q*uality-cost trade-offs*

Whether we are speaking of a competitive system or one dominated by the public sector, the link between performance and financing remains largely informal. There are two reasons why this should not surprise us. On one hand, despite the progress in performance measurement studies, hospital activity is so varied that it is difficult to summarize with a few indicators. The relationship between the level of these indicators and the cost of care has yet to be thoroughly examined. Study of this relationship is called for, using benchmarking methods among others. If such benchmarking is to be effective, hospital cooperation is indispensable. Incentives must therefore be found to encourage such cooperation, by sharing the risks and benefits of efforts to provide a better understanding of hospital performance. On the other hand, formalizing this price-quality relationship will require purchasers to be explicit about the price they are ready to pay for an additional unit of quality. A purchaser operating in a competitive environment should in theory obtain this information from the market, while a public purchaser should in principle rely on democratic debate in setting its price. Yet revealing popular preferences is a very difficult task for any government, whether in the health field or in other areas of public policy.

© OECD 2002

NOTES

1. The sole clinical outcome indicator in Dr. Foster is mortality. The Web site also publishes a descriptive fact sheet for each hospital, comparative outcomes focused primarily on waiting times for each specialty, the number of complaints and the number of complaints processed.
2. In order to compare the performance of health plans, the NCQA has developed a comparative database known as HEDIS (Health Plan Employer Data and Information Set) to which NCQA-accredited plans contribute on a voluntary basis.
3. For the sake of style, we shall use here the terms performance and quality of care interchangeably, to designate, on one hand, indicators that focus on outcomes measured on a biological or clinical basis, and on the other hand the level achieved by a hospital against these indicators. We are not speaking here of efficiency measurement, which would also take into account patient need satisfaction in other dimensions, compared to the costs involved.
4. It may also be that quality preferences are relatively insensitive to the client-perceived level of risk. In France, for example, a significant obstacle to introducing a maternity system ranked in terms of its ability to treat high-risk pregnancies is that low-risk women, who could comfortably give birth in "level 1" facilities, nevertheless prefer the security offered by the high-tech "level 3" maternity hospitals, even when they do not need such care.

© OECD 2002

REFERENCES

ANSARI, M.Z., ACKLAND, M.J., JOLLEY, D.J., CARSON, N. and MCDONALD, I.G. (1999), "Inter-hospital comparison of mortality rates", *International Journal of Quality in Health Care*, Vol. 2, No. 1, pp. 29-35.

ARON, D.C., HARMER, D.L., SHEPARDSON, L.B. and ROSENTHAL, G.E. (1998), "Impact of risk-adjusting Cesarean delivery rates when reporting hospital performance", JAMA, Vol. 279, No. 24, pp. 1968-1972.

BUCKNALL, C.E., RYLAND, I., COOPER, A., COUTTS, H., CONNOLLY, C.K. and PEARSON, M.G. (2000), "National benchmarking as a support system for clinical governance", *Journal of the Royal College of Physicians of London*, Vol. 34, No. 1, pp. 52-56.

CARON, A. and NEUHAUSER, D. (1999), "The effect of public accountability on hospital performance: trends in rates for Cesarean section rates and vaginal births after Cesarean section in Cleveland, Ohio", *Quality Management in Health Care*, Vol. 7, No. 2, pp. 1-10.

CHALÉ, J.J. and NAIDITCH, M. (1999), "Retour sur une enquête controversée", *La Recherche*, No. 324, pp. 71-74.

CHASSIN, M.R., HANNAN, E.L. and DE BUONO, B.A. (1996), "Benefits and hazards of reporting medical outcomes publicly", NEJM, Vol. 334, No. 6, pp. 394-398.

CIHI (2001), *Hospital Report* 2001. *www.cihi.ca/HospitalReport/*

COMAROW, A. (2000*a*), "Higher volumes, fewer deaths", USN*ews*, July 17.

COMAROW, A. (2000*b*), "Picking the top centers. Reputation, outcomes and technology all count", USN*ews*, July 17.

COOPER, J.K., KOHLMANN, T., MICHAEL, J.A., HAFFER, S.C. and STEVIC, M. (2001), "Health outcomes. New quality measure for Medicare", *International Journal for Quality in Health Care*, Vol. 13, No. 1, pp. 9-16.

COULTER, A. and CLEARY, P. (2001), "Patients' experience with hospital care in five countries", *Health Affairs*, Vol. 20, No. 3, pp. 244-252.

DREES (1999), "Les éléments de la 'performance' hospitalière", *Études et Résultats*, No. 42, Paris.

DREES (2001), "Eléments pour évaluer la performance des établissements hospitaliers", *Dossiers solidarité et santé*, No. 2, Paris.

EDDY, D.M. (1998), "Performance measurement: problems and solutions", *Health Affairs*, Vol. 17, No. 4, pp. 7-25.

EPSTEIN, M. and KURTZIG, B. (1994), "Statewide health information: a tool for improving hospital accountability", *The Joint Commission Journal of Quality Improvement*, Vol. 20, No. 7, pp. 370-375.

HANNAN, E.L., KILBURN, J.F., O'DONNEL, J.F. *et al.* (1990), "Adult open heart surgery in New York State. An analysis of risk factors and hospital mortality rates", JAMA, pp. 2768-2774.

HANNAN, E.L., KILBURN, J.F., RACZ, M. *et al.* (1994), "Improving the outcomes of Coronary Bypass Surgery in New York State", JAMA, No. 271, pp. 761-766.

HANNAN, E.L., SIU, A.L., KUMAR, D. *et al.* (1995), "The decline in Coronary Bypass Graft Surgery Mortality in New York State. The role of surgeon volume", JAMA, No. 273, pp. 209-213.

HIBBARD, J.H., SLOVIC, P. and JEWETT, J.J. (1997), "Informing consumer decisions in health care: implications from decision-making research", *The Milbank Quarterly*, Vol. 75, No. 3, pp. 395-414.

© OECD 2002

HIBBARD, J.H., SOFAER, S. and JEWETT, J.J. (1996),
"Condition-specific performance information: assessing salience, comprehension, and approaches for communicating quality", *Healthcare Financing Review*, Vol. 18, No. 1, pp. 95-111.

HILLNER, B.E., SMITH, T.J. and DESCH, C.E. (2000),
"Hospital and physician volume or specialization and outcomes in cancer treatment: importance in quality of cancer care", *J Clin Oncol*, Vol. 18, No. 11, pp. 2327-2340.

HUMPHREYS, H. and EMMERSON, A.M. (1993),
"Control of hospital acquired infection: accurate data and more resources, not league tables", *Journal of Hospital Infection*, No. 25, pp. 75-78.

IEZZONI, L. (1997),
"The risks of risk adjustment", JAMA, Vol. 278, No. 19, pp. 1600-1607.

IEZZONI, L., SCHWARTZ, M., ASH, A.S., HUGHES, J.S., DALEY, J. and MACKIERNAN, Y.D. (1995),
"Using severity-adjusted stroke mortality rates to judge hospitals", *International Journal for Quality in Health Care*, Vol. 7, No. 2, pp. 81-94.

JEWETT, J.J. and HIBBARD, J.H. (1996),
"Comprehension of quality indicators: differences among privately insured, publicly insured and uninsured", *Healthcare Financing Review*, Vol. 18, No. 1, pp. 75-94.

Joint Commission on Accreditation of Health Care Organizations (2001),
The Self-assessment Checklist: Hospitals, Oakbrook terrace, Illinois, United States.

KRITSCHESKY, S.B., BRAUN, B.I., GROSS, P.A., NEWCOMB, C.S., KELLEHER, C.A. and SIMMONS, B.P. (1999),
"Definition and adjustment of Cesarean section rates and assessments of hospital performance", *International Journal for Quality in Health Care*, Vol. 2, No. 4, pp. 283-291.

LANDON, B., IEZZONI, L., ASH, A., SHWARTZ, M., DALEY, J., HUGUES, J.S. and MACKIERNAN, Y.D. (1996),
"Judging hospitals by severity adjusted mortality rates: the case of CABG surgery", *Inquiry*, No. 33, pp. 155-166.

LEYLAND, A.H. and BODDY, F.A. (1998),
"League tables and acute myocardial infarction", *The Lancet*, No. 351, pp. 555-558.

LIANG, B.A. (2000),
"Risks of reporting sentinel events", *Health Affairs*, Vol. 19, No. 5, pp. 113-120.

LOMBRAIL, P. and NAIDITCH, M. (1999),
"Comment partager l'information ?", *La Recherche*, No. 324, pp. 75-78.

MARSHALL, R. (2001),
"Where We Get the Best Health Care", *Macleans*, No. 114, pp.31-36.

MCCORMACK, L.A., GARFINKEL, S.A., SCHNAIER, J.A., LEE, A.J. and SANGL, J.A. (1996),
"Consumer information development and use", *Healthcare Financing Review*, Vol. 18, No. 1, pp. 15-31.

MCKEE, M. and HUNTER, D. (1995),
"Hospital league tables: do they inform or mislead?", *Quality in Health Care*, No. 4, pp. 5-12.

MCLAUGHLIN, C.G. and GINSBURG, P.B. (1998),
"Competition, quality of care and the role of the consumer", *The Milbank Quarterly*, Vol. 76, No. 4, pp. 737-743.

MC MULLAN, M. (1996),
"HCFA's consumer information commitment", *Healthcare Financing Review*, Vol. 18, No. 1, pp. 9-15.

MINVIELLE, E. (1996),
"Gérer la singularité à grande échelle", *Revue Française de Gestion*, No. 109, pp. 119-124.

NHS (1995),
The NHS Performance Guide 1994-1995, London.

NAIDITCH, M. (1999),
"L'exemple de la chirurgie cardiaque new-yorkaise", *La Recherche*, No. 324, pp. 66-70.

NATHENS, A.B., JURKOVICH, G.J., GROSSMAN, D.C., MACKENZIE, E.J. and RIVARA, F.P. (2001),
"Relationship between trauma center volume and outcomes", JAMA, Vol. 285, No. 9, pp. 1164-1171.

PHILLIPS, K.A., LUFT, H.S. and Ritchie, J.L. (1995),
"The association of hospital volumes of percutaneous transluminal coronary angioplasty with adverse outcomes, length of stay, and charges in California", *Medical Care*, Vol. 33, No. 5, pp. 502-514.

LE POINT (2001),
"Hôpitaux. Le Palmarès 2001", No. 1151, August 31.

RITCHIE, J.L., MAYNARD, C., CHAPKO, M.K., EVERY, N.R. and MARTIN, D.C. (1999),
"Association between percutaneous transluminal coronary angioplasty volumes and outcomes in the Health Care Cost and Utilization Project 1993-1994", AM J *Cardiol*, Vol. 83, No. 4, pp. 493-497.

© OECD 2002

ROMANO, P.S., ZACH, A., LUFT, H.S., RAINWATER, J., RELY, L.L. and CAMPA, D. (1995),
"The California Hospital Outcomes Project: using administrative data to compare hospital performance", *The Joint Commission Journal on Quality Improvement*, Vol. 21, No. 12, pp. 668-685.

ROSENTHAL, G. (1997),
"Weak association between hospital mortality for individual diagnoses: implications for profiling hospital quality", *American Journal of Public Health*, Vol. 87, No. 3, pp. 429-433.

ROSENTHAL, G.E., QUINN, L. and HARPER, D.L. (1997),
"Declines in Hospital Mortality associated with a regional initiative to measure hospital performance", *American Journal of Medical Quality*, Vol. 12, No. 2, pp. 103-112.

ROSENTHAL, G.E., HAMMAR, P.J., WAY, L.E., SHIPLEY, S.A., DONER, D., WOJTALA, B., MILLER, J. and HARMER, D.L. (1998),
"Using hospital performance data in quality improvement: the Cleveland Health Quality Choice Experience", *The Joint Commission Journal on Quality Improvement*, Vol. 24, No. 7, pp. 347-360.

SHAW, C.D. (1997),
"Health-care league tables in the United Kingdom", *J. Qual. Clin. Practice*, No. 17, pp. 215-219.

SILBER, J.H. and ROSENBAUM, P.R. (1997),
"A spurious correlation between hospital mortality and complication rates: the importance of severity adjustment", *Medical Care*, No. 35(suppl.), pp. OS77-OS92.

SIRIO, C.A., ANGUS, D.C. and ROSENTHAL, G.E. (1994),
"Cleveland Health Quality Choice (CHQC) – An ongoing collaborative, community based outcomes assessment program", *New Horizons*, Vol. 2, No. 3, pp. 321-325.

SISK, J. (1998),
"Increased competition and the quality of health care", *The Milbank Quarterly*, Vol. 76, No. 4, pp. 687-707.

THOMAS, J.W. and HOFER, T.P. (1999),
"Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care", *Medical Care*, Vol. 37, No. 1, pp. 83-92.

VOELKER, R. (1995),
"Why has historic' diclosure of hospital performance data attracted so little attention?", JAMA, Vol. 273, No. 9, pp. 689-690.

© OECD 2002

TABLE OF CONTENTS

Editor's Preface
by *Peter Smith* 7

Part I
Overview of Issues and Challenges

Chapter 1. Measuring Health System Performance: Problems and Opportunities in the Era of Assessment and Accountability
by *C. David Naylor, Karey Iron* and *Kiren Handa* 13

Chapter 2. Performance Measurement and Improvement in OECD Health Systems: Overview of Issues and Challenges
by *Jeremy Hurst* 35

Part II
Performance Measurement and Performance Management from the Perspective of Various Actors: A Review of Experiences in Selected Countries

Chapter 3. Improving Value for Money in the United Kingdom National Health Service: Performance Measurement and Improvement in a Centralised System
by *Clive H. Smee* 57

Chapter 4. Improving Accountability in a Decentralised System: A Swedish Perspective
by *Nina Rehnqvist* 87

Chapter 5. Providing Performance Information for Consumers: Experience from the United States
by *Thomas Reilly, Gregg Meyer, Carla Zema, Christine Crofton, David Larson, Charles Darby* and *Katherine Crosson* 97

Chapter 6. Can a Tulip Become a Rose? The Dutch Route of Guided Self-Regulation Towards a Community-Based Integrated Health Care System
by *Niek Klazinga, Diana Delnoij* and *Isik Kulu-Glasgow* 117

Chapter 7. Towards Integrated and Coherent Health Information Systems for Performance Monitoring: The Canadian Experience
by *Michael Wolfson* and *Richard Alvarez* 133

Part III
Performance Measurement Activities at the International Level: How can International Comparisons Assist National Policy-Making?

Chapter 8. Opening the Black Box: What Can Be Learned from a Disease-based Approach?
Stephane Jacobzone, Pierre Moise and *Lynelle Moon* 159

© OECD 2002

Chapter 9. The Evolution of WHO's Approach to Health System Performance Assessment
by *David* B. *Evans* 197

Chapter 10. Measuring and Improving Patients' Experiences: How Can we Make Health Care Systems Work for Patients?
by *Angela Coulter* and *Paul Cleary* 211

Chapter 11. Equity in the Use of Physician Visits in OECD Countries: Has Equal Treatment for Equal Need Been Achieved?
by *Eddy van Doorslaer*, *Xander Koolman* and *Frank Puffer* 225

Part IV
Best Practices in Measuring Different Dimensions of Health System Performance

Chapter 12. Measuring the Quality of Hospital Care: The State of the Art
by *Gérard de Pouvourville* and *Étienne Minvielle* 251

Chapter 13. Measuring the Quality of Long-term Care in Institutional and Community Settings
by *Naoki Ikegami*, *John* P. *Hirdes* and *Iain Carpenter* 277

Chapter 14. Developing Composite Indicators for Assessing Health System Efficiency
by *Peter Smith* 295

Part V
Applying Performance Indicators to Health System Improvement

Chapter 15. Applying Performance Indicators to Health System Improvement
by *Sheila Leatherman* 319

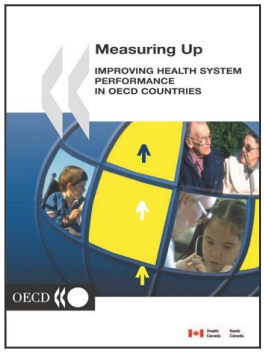
Part VI
Summary and Conclusions

Chapter 16. Measuring Up: Lessons and Potential
by *Jean-François Girard* and *Étienne Minvielle* 337

Part VII
Ministerial Roundtable: Leadership, Successes and Challenges

Chapter 17. Measuring Health System Performance and the Impact on Political Decision-Making: The Views of OECD Ministers
by *Julio Frenk* (*Minister of Health, Mexico*), *John Hutton* (*Minister of State for Health, United Kingdom*), *Bernard Kouchner* (*Minister of Health, France*), *Allan Rock* (*Minister of Health, Canada*), *Edward Sondik* (*on behalf of Tommy Thompson, Secretary of Health* and *Human Services, United States*) 351

© OECD 2002



From:

Measuring Up

Improving Health System Performance in OECD Countries

Access the complete publication at:

https://doi.org/10.1787/9789264195950-en

Please cite this chapter as:

OECD (2003), "Measuring the Quality of Hospital Care: The State of the Art", in *Measuring Up: Improving Health System Performance in OECD Countries*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264195950-14-en

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. Extracts from publications may be subject to additional disclaimers, which are set out in the complete version of the publication, available at the link provided.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at http://www.oecd.org/termsandconditions.