



WHO European Ministerial
Conference on Health Systems:
"HEALTH SYSTEMS,
HEALTH AND WEALTH"

Tallinn, Estonia, 25–27 June 2008

BACKGROUND DOCUMENT

Performance measurement for health system improvement: experiences, challenges and prospects

Peter C. Smith, Elias Mossialos and
Irene Papanicolas

Performance measurement for health system improvement: experiences, challenges and prospects

Peter C. Smith, Elias Mossialos and Irene Papanicolas

Keywords:

DELIVERY OF HEALTH CARE - standards

QUALITY INDICATORS, HEALTH CARE

QUALITY ASSURANCE, HEALTH CARE - organization and administration

EUROPE

© World Health Organization 2008 and World Health Organization, on behalf of the European Observatory on Health Systems and Policies 2008

Address requests about publications of the WHO Regional Office for Europe to:

Publications
WHO Regional Office for Europe
Scherfigsvej 8
DK-2100 Copenhagen Ø, Denmark

Alternatively, complete an online request form for documentation, health information, or for permission to quote or translate, on the Regional Office web site (<http://www.euro.who.int/pubrequest>).

All rights reserved. The Regional Office for Europe of the World Health Organization welcomes requests for permission to reproduce or translate its publications, in part or in full.

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by the World Health Organization to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either express or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall the World Health Organization be liable for damages arising from its use. The views expressed by authors, editors, or expert groups do not necessarily represent the decisions or the stated policy of the World Health Organization.

Contents

	Page
Key messages	i
Executive summary	ii
Performance measurement for health system improvement: experiences, challenges and prospects	
Policy issue	1
Purpose of performance measurement	1
Defining and measuring performance	2
Methodological issues about performance measurement	6
Using performance measurement: key policy levers	9
Summary and conclusions	15
References	18

Authors

Elias Mossialos, LSE Health and European Observatory on Health Systems and Policies, London School of Economics and Political Science, United Kingdom

Irene Papanicolas, LSE Health, London School of Economics and Political Science, United Kingdom

Peter C. Smith, Centre for Health Economics, University of York, United Kingdom

Key messages

- Performance measurement offers policy-makers a major opportunity for securing health system improvement and accountability.
- Performance measurement aims to improve the quality of decisions made by all actors within the health system.
- Securing improved performance measurement often requires the active leadership of government.
- Major improvements are still needed in data collection, analytical methodologies, policy development and implementation of health system performance measurement.
- Definitions of performance indicators should be clear, consistent and fit into a clear conceptual framework.
- Policy-makers should pay particular attention to the political and organizational context within which performance data are collected and disseminated.
- Considerable progress has been made in developing performance indicators for acute hospital care, primary care and population health, but for mental health, financial protection and health system responsiveness research is at a much earlier stage of development.
- The development of individual performance indicators requires concerted expert and political attention, and these indicators should: aim to provide information that is relevant to the needs of specific actors; attempt to measure performance that is directly attributable to an organization or actor; aim to be statistically sound, easy to interpret and unambiguous; and be presented with full acknowledgement of any data limitations.
- The presentation of performance measurement data and how this influences its interpretation by patients, providers and practitioners and the public require more attention.
- Public reporting has many benefits, but can lead to adverse outcomes; mechanisms should be put in place to monitor and counteract these adverse outcomes.
- An important use of performance measurement is to provide feedback to clinical practitioners on their actions and how these compare to those of their peers.
- Performance measurement systems should be monitored frequently to ensure alignment with other health system mechanisms and to identify areas for improvement.
- Experiments under way to examine how performance measurement can be used in conjunction with explicit financial incentives to reward provider performance are a promising area for policy and a priority for further research.
- A better evidence base on which to underpin performance measurement policy is needed, and new initiatives should be subject to rigorous evaluation.

Executive summary

Performance measurement offers policy-makers a major opportunity to secure health system improvement and accountability. Its role is to improve the quality of decisions made by all actors within the health system, including patients, practitioners, managers, governments at all levels, insurers and other payers, politicians, and citizens as financial supporters.

Recent major advances in information technology and increasing demands for health system accountability and patient choice have driven rapid advances in health system performance measurement. Health systems, however, are still in the relatively early stages of performance measurement, and major improvements are still needed in data collection, analytical methodologies, and policy development and implementation.

Health system performance has a number of aspects – including population health, health outcomes from treatment, clinical quality and the appropriateness of care, responsiveness, equity and productivity – and progress is varied in the development of performance measures and data collection techniques for these different aspects. Considerable progress has been made in such areas as acute hospital care, primary care and population health, but in such areas as mental health, financial protection and health system responsiveness, research is at a much earlier stage of development.

The first requirement of any performance measurement system is to formulate a robust conceptual framework within which performance measures can be developed. Definitions of performance indicators should then fit into the framework and satisfy a number of criteria, such as face validity, reproducibility, acceptability, feasibility, reliability, sensitivity and predictive validity. Besides paying attention to these technical considerations, policy-makers should pay careful attention to the political and organizational context within which performance data are to be collected and disseminated.

Numerous technical questions arise when analysing and interpreting performance measures. Among the most important are: what has caused the observed performance and to what practitioners, organizations or agencies should variations in performance be attributed? In some areas, advanced analytical methods of risk adjustment have been developed to help answer the question about attribution.

In some aspects of health care, patient safety is a major concern, and methods of statistical surveillance have been developed to help detect anomalous performance rapidly and confidently. An example of anomalous, though not necessarily unsafe, performance is the

overuse of a particular intervention, and the need to find out whether it means something unsafe for patients would then follow the initial finding of an *anomaly*.

More attention should be paid to the presentation of performance-measurement data and how patients, providers, practitioners and the public interpret it and are influenced by it. For example, a particularly contentious issue is the use of *composite* measures of performance, which seek to combine several performance indicators into a single measure of organizational or system performance. These are superficially attractive, as they can help summarize levels of attainment in an accessible fashion, but they can also lead to faulty inferences and should be used with caution.

Policy-makers can use performance measurement in a number of ways to promote system improvement. It can be used in *public reporting of performance*, sometimes in the form of organizational *report cards*. This has been found to have an important beneficial effect, particularly on provider organizations. However, it has so far had little direct effect on patients and can also lead to adverse outcomes, such as avoidance of patients with complex health problems. Mechanisms should be put in place to monitor and counteract such tendencies.

Experiments are under way to examine how performance measurement can be used with *explicit financial incentives* to reward health care provider performance. This is a promising policy area. However, such schemes raise a number of important questions about design, such as which aspects of performance to target, how to measure attainment, how to set targets, whether to offer incentives at the individual or group level, how strong to make the link between achievement and reward, and how much money to attach to an incentive. So far, there is little convincing research evidence of the effectiveness of such incentives, and this is a priority for further research.

Targets, a quantitative expression of an objective to be met in the future, are a particular form of incentive mechanism. They have been particularly prevalent in the area of public health. Their effectiveness in securing major system improvements, however, has been questioned, and it is unlikely that they will secure such improvements unless aligned with other policy levers, such as strong democratic accountability, market mechanisms or direct financial incentives.

Performance measurement can also be used to provide *feedback to clinical practitioners* on their performance relative to their peers. These feedback systems can secure widespread improvements in performance. However, to be successful, they need to be owned by

the practitioners and usually require careful statistical risk adjustment to control for confounding patient characteristics. Also, the need to provide feedback that does not immediately threaten the reputation or livelihood of clinicians and other professionals can at times conflict with the demand for public reporting.

Securing improved performance measurement is an important stewardship task of government, as many of the benefits of performance measurement cannot be realized without the active leadership of government, whether through law, regulation, coordination or persuasion. Stewardship responsibilities associated with performance measurement can be summarized under the following headings:

1. *development of a clear conceptual framework and a clear vision of the purpose of the performance measurement system:*
 - alignment with accountability relationships;
 - alignment with other health system mechanisms, such as finance, market structure and information technology;
2. *design of data collection mechanisms:*
 - detailed specification of individual indicators;
 - alignment with international best practice;
3. *information governance:*
 - data audit and quality control;
 - ensuring public trust in information;
 - ensuring well-informed public debate;
4. *development of analytical devices and capacity to help understand the data:*
 - ensuring analysis is undertaken efficiently and effectively;
 - ensuring local decision-makers understand the analysis;
 - commissioning appropriate research on, for example, risk adjustment, uncertainty and data feedback mechanisms;
5. *development of appropriate data aggregation and presentational methods:*
 - ensuring information has appropriate effect on all parties;
 - mandating public release of summary comparative information;
 - ensuring comparability and consistency;

6. *design of incentives to act on performance measures:*

- monitoring effect of performance information on behaviour;
- acting to enhance beneficial outcomes and negate any adverse consequences;

7. *proper evaluation of performance-measurement instruments:*

- ensuring money is spent cost-effectively on information resources;

8. *managing the political process:*

- developing and monitoring policy options;
- encouraging healthy political debate;
- ensuring that specific interest groups do not capture the performance information system.

None of these roles need be undertaken by government itself, but it must be ensured that they all function effectively.

Performance measurement for health system improvement: experiences, challenges and prospects

Policy issue

Information plays a central role in the ability of a health system to secure improved health effectively and efficiently for its population. It can be used in many diverse ways, such as tracking public health, monitoring health care safety, determining appropriate treatment paths for patients, promoting professional improvement, ensuring managerial control and promoting the accountability of the health system to the public. Underlying all of these efforts is the role performance measurement plays in guiding the decisions that various stakeholders – such as patients, clinicians, managers, governments and the public – make in steering the health system towards better outcomes.

Records of performance-measurement efforts in health systems can be traced back at least 250 years (1,2). More formal arguments for the collection and publication of information on performance were developed more than 100 years ago, when such pioneers in the field as Florence Nightingale and Ernest Codman campaigned for its widespread use in health care. Until recently, professional, practical, and political barriers have prevented these principles from becoming a reality (3). For example, Nightingale's and Codman's efforts were frustrated by professional resistance and, until recently, information systems have failed to deliver their promised benefits, in the form of timely, accurate and comprehensive information.

Over the past 25 years, however, health system performance measurement and reporting have grown substantially, thus helping to secure health system improvement. Many factors have contributed to this growth. On the demand side, health systems have come under intense pressure to contain costs; also, patients now expect to make more informed decisions about their treatment, and strong demands have been made for increased audit and accountability of the health professions and health service institutions (4,5). On the supply side, the great advances in information technology have made it much cheaper and easier to collect, process and disseminate data.

In many respects, the policy agenda is moving away from discussions of whether performance measurement should be undertaken and what data to collect and is moving towards determining the best ways in which to summarize and present such data and how to integrate it successfully into effective structures for governance. Yet, despite the proliferation of performance-measurement initiatives, there remain a large number of unresolved questions about the collection and deployment of such information. Health systems are still experimenting with the concept of performance

measurement, and much still needs to be done to realize its full potential.

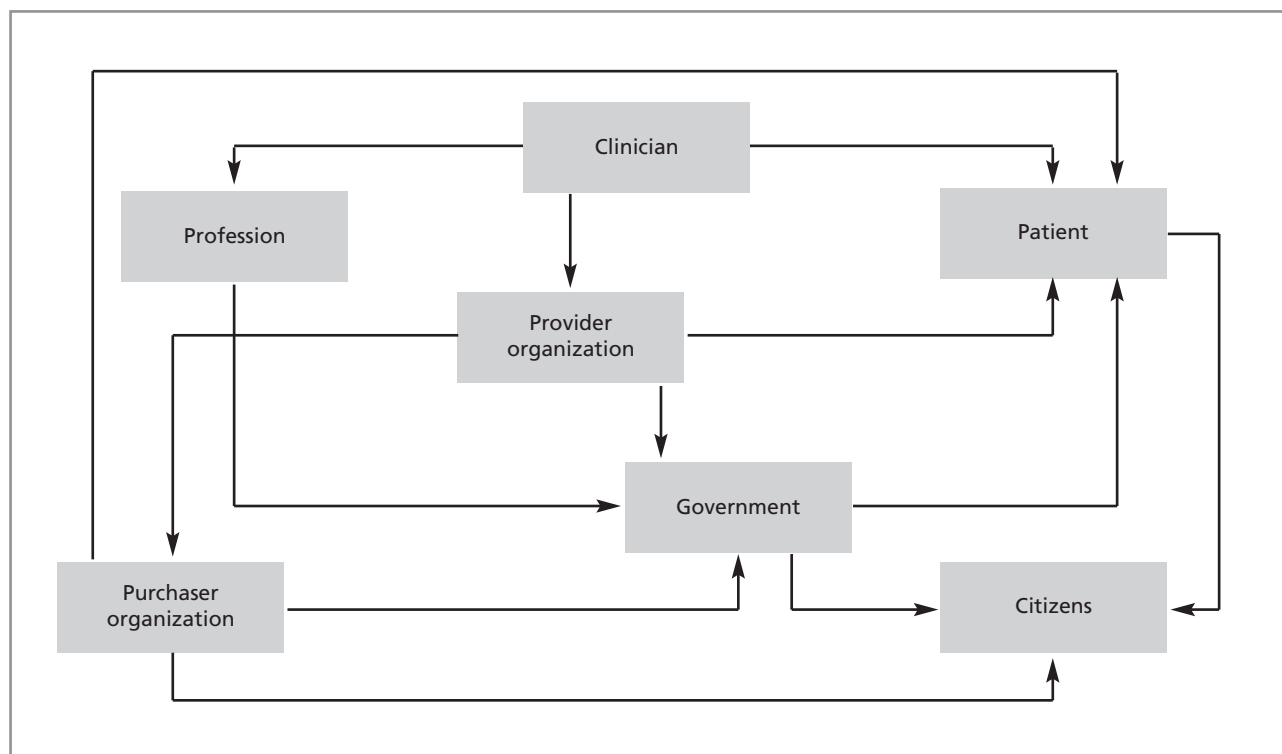
This document reviews some of the main issues emerging in the debate about performance measurement and draws on a detailed collection of essays by leading experts in the field. These essays were prepared for the WHO Ministerial Conference on Health Systems, in Tallinn, Estonia, and are due to be published after the conference by Cambridge University Press (6). The document first examines the purpose of performance measurement and the different areas for which data are collected. It then examines the different ways in which performance measurement has been presented and used for health system improvement internationally. Finally, it discusses the major challenges found in presenting and using performance measures and concludes by presenting key lessons and future priorities for policy-makers.

Purpose of performance measurement

Health systems are complex entities with many different stakeholders, including patients, clinicians, health care providers, purchaser organizations, regulators, the government and the broader public. These stakeholders are linked by a series of accountability relationships (Fig. 1). Accountability has two broad elements: the rendering of an account (providing information) and the consequent holding to account (sanctions or rewards for the accountable party). Whatever the precise design of the health system, the fundamental role of performance measurement is to help hold its various agents to account, by enabling stakeholders to make informed decisions. It is therefore noteworthy that, if the accountability relationships are to function properly, no system of performance information should be viewed in isolation from the broader system design within which the measurement is embedded.

Each of the relationships described in Fig. 1 has different information needs in terms of the nature of the information, its detail and timeliness, and the level of aggregation required. For example, in choosing which provider to use, a patient may need detailed comparative data on health outcomes. In contrast, in holding a government to account and in deciding for whom to vote, a citizen may need highly aggregated summaries and trends. Many intermediate needs also arise. In deciding whether providers are performing adequately, a purchaser (such as a social insurer) may need both broad, more aggregated information and detailed assurance of safety aspects. A fundamental challenge for performance measurement is to design information systems that serve these diverse needs. Table 1 examines this issue in more detail.

Fig. 1. A map of some important accountability relationships in the health system



In practice, the development of performance measurement has rarely been pursued with a clear picture of who the information users are or what their information needs might be. Instead performance-measurement systems have usually sought to inform a variety of users, typically presenting a wide range of data in the hope that some of the information collected will be useful to different parties. Yet, given the diverse information needs of the different stakeholders in health systems, it is unlikely that a single method of reporting performance will be useful for everybody. Instead, data sources should be designed and exploited to satisfy the demands of different users. This may often involve using data from the same sources in different forms. A major challenge for health systems is, therefore, to develop more nuances in the collection and presentation of performance measures for the different stakeholders without imposing a huge burden of new data collection and analysis.

Defining and measuring performance

In general, performance measurement seeks to monitor, evaluate and communicate the extent to which various aspects of the health system meet their key objectives. Usually, those objectives can be summarized under a limited number of headings – for example, health conferred on people by the health system, its responsiveness to public preferences, the financial protection it offers and its productivity. *Health* relates

both to the health outcomes secured after treatment and to the broader health status of the population. *Responsiveness* captures aspects of health system behaviour not directly related to health outcomes, such as dignity, communication, autonomy, prompt service, access to social support during care, quality of basic services and choice of provider. *Productivity* refers to the extent to which the resources used by the health system are used efficiently in the pursuit of effectiveness. Besides a concern for the overall attainment in each of these areas, *The world health report 2000 (7)* highlighted the importance of distributional (or equity) issues, expressed in terms of inequity in health outcomes, responsiveness and payment. Table 2 summarizes these largely universal aspects of health performance measures.

Various degrees of progress have been made in the development of performance measures and data collection techniques for the different aspects of health performance. In some areas, such as population health, there are well-established indicators – for example, infant mortality and life expectancy (sometimes adjusted for disability). Even here, however, important further work is needed. A particular difficulty with population health measures is estimating the specific contribution of the health system to health. To address this, researchers are developing new instruments, such as the concept of *avoidable mortality (8,9)*.

Table 1. Information requirements for stakeholders in health care systems

Stakeholder	Examples of needs	Data requirements
Government	Monitoring the health of the nation Setting health policy Ensuring that regulatory procedures are working properly Ensuring that government finances are used as intended Ensuring that appropriate information and research functions are undertaken Monitoring regulatory effectiveness and efficiency	Information on performance at national and international levels Information on access and equity of care Information on utilization of service and waiting times Population health data
Regulators	Protecting patient's safety and welfare Ensuring broader consumer protection Ensuring the market is functioning efficiently	Timely, reliable and continuous information on patient safety and welfare Information on probity and efficiency of financial flows
Payers (taxpayers and members of insurance funds)	Ensuring money is being spent effectively, efficiently and in line with expectations	Aggregate, comparative performance measures Information on productivity and cost-effectiveness Information on access to (and equity of) care
Purchaser organizations	Ensuring that contracts offered to their patients are in line with the objectives the patients expect	Information on patient experiences and patient satisfaction Information on provider performance Information on the cost-effectiveness of treatments
Provider organizations	Monitoring and improving existing services Assessing local needs	Aggregated clinical performance data Information on patient experiences and patient satisfaction Information on access and equity of care Information on utilization of service and waiting times
Physicians	Staying up to date with current practice Being able to improve performance	Information on current practice and best practice Performance information benchmarks
Patients	Being able to make a choice of provider when in need Information on alternative treatments	Information on location and quality of nearby emergency health services Information on quality of options for elective care
The public	Being reassured that appropriate services will be available if needed in the future Holding government and other elected officials to account	Broad trends in and comparisons of system performance at national and local level Efficiency information Safety information

The contribution of the health system to health care can be more reliably captured in terms of clinical outcomes for patients. Traditionally, this contribution has been examined using post-treatment mortality, which is a blunt instrument. However, increasing interest is focusing on more general measures of improvements in patient health status, often in the form of patient-

reported outcome measures. These measures are derived from simple surveys of subjective health status administered directly to patients, often before and after treatment. Numerous instruments have been developed, often in the context of clinical trials. These take the form of detailed condition-specific questionnaires or broad-brush generic measures (10).

Table 2. Aspects of health performance measures

Measurement area	Description of measures	Examples of indicators
Population health	Measures of aggregated data on the health of the population	Life expectancy Years of life lost Avoidable mortality Disability-adjusted life-years
Individual health outcomes	Measures of individual's health status, which can be relative to the whole population or among groups Indicators that also apply utility rankings to different health states	Generic measures: • Short form 36 (SF-36) ^a • EQ-5D ^b Disease-specific measures: • arthritis impact measurement scale • Parkinson's disease questionnaire (PDQ-39)
Clinical quality and appropriateness of care	Measures of the services and care patients receive to achieve desired outcomes Measures used to determine if best practice takes place and whether these actions are carried out in a technologically sound manner	Outcome measures: • health status • specific post-operative readmission and mortality rates Process measures: • frequency of blood pressure measurement
Responsiveness of health system	Measures of the way individuals are treated and the environment in which they are treated during health system interactions Measures concerned with issues of patient dignity, autonomy, confidentiality, communication, prompt attention, social support and quality of basic amenities	Patient experience measures Patient satisfaction measures
Equity	Measures of the extent to which there is equity in health, access to health care, responsiveness and financing	Utilization measures Rates of access Use-needs ratios Spending thresholds Disaggregated health outcome measures
Productivity	Measures of the productivity of the health care system, health care organizations and individual practitioners	Labour productivity Cost-effectiveness measures (for interventions) Technical efficiency (measures of output/input) Allocative efficiency (measured by willingness to pay)

^a SF-36 is a multipurpose, short-form health survey with only 36 questions.

^b EQ-5D is a standardized instrument for measuring the outcome of a wide range of health conditions and treatments. It provides a simple descriptive profile and a single index value for health status that can be used in the clinical and economic evaluation of health care and in population health surveys.

Source: Smith et al. (6).

To measure performance when monitoring outcomes from health care interventions over time and between providers, the policy challenge is to identify the most appropriate choice of instrument. In England, for example, the government has recently mandated the use of the generic patient-reported outcome measure instrument EQ-5D for use for all National Health Service patients undergoing four common procedures. This experiment will assess the costs of such routine use and will test whether the resistance of some health

professionals to patient-reported outcome measures is sustained. Also, while the relevance of patient-reported outcome measures to acute care is clear, their application to such areas as chronic disease and mental illness remain less well developed.

Although clinical outcome measures are the gold standard for measuring effectiveness in health care, their use can be problematic – for example, if the outcomes cannot realistically be assessed in a timely or feasible

Table 3. Usefulness of structural outcome and process indicators

Type of indicator	Advantages	Disadvantages	Most useful areas
Outcome	<ul style="list-style-type: none"> Often more meaningful to stakeholders Attention directed to (and health goals focused on) the patient Encourages long-term health-promotion strategies Not easily manipulated 	<ul style="list-style-type: none"> May be ambiguous and difficult to interpret, as they are the result of many factors that are difficult to disentangle Takes time to collect Requires a large sample size to detect statistically significant effects Can be difficult to measure – for example, wound infection 	<ul style="list-style-type: none"> To measure quality of homogeneous procedures To measure quality of homogeneous diagnosis with strong links between interventions and outcomes To measure quality of interventions made to heterogeneous populations that suffer from a common condition
Process	<ul style="list-style-type: none"> Easily measured without major bias or error More sensitive to quality of care Easier to interpret Require a smaller sample size to detect statistically significant effects Can often be observed unobtrusively Provide clear pathways for action Capture aspects of care valued by patients (aside from outcomes) 	<ul style="list-style-type: none"> Often too specific, focusing on a particular intervention or condition May quickly become dated as models of care and technology develop May have little value to patients unless they understand how they relate to outcomes May be easily manipulated 	<ul style="list-style-type: none"> To measure quality of care, especially for treatments where technical skill is relatively unimportant To measure quality of care of the homogeneous conditions in different settings

Source: Adapted from Davies (13) and Mant (14).

fashion. This is particularly important for chronic diseases. Measures of the process then become important signals of future success (11). Process measures are based on actions or structures known to be associated with health system outcomes, in either health or responsiveness. An example of an action might be appropriate prescribing, which is known from research evidence to contribute to good outcomes (12). Also, the concept of *effective coverage* is an important population health process measure. Table 3 summarizes the basic advantages and disadvantages of using outcome and process indicators and the areas of performance measurement where they are most useful.

Work in the area of responsiveness is inherently challenging, as in principle it requires general surveys of both users and non-users of health services. Also, aggregating diverse areas into usable summary indicators of responsiveness is problematic. The World Health Survey of households in over 70 countries contained a *responsiveness* module that offers some potential for proposing operational solutions to the routine measurement of health system responsiveness (15).

Financial protection from the catastrophic expenditure associated with ill health is a fundamental health system

concern. Many high-income countries have introduced universal insurance coverage to address this issue, but even then there are quite large variations in measures of financial protection between countries and over time. The issue, however, is even more acute in many lower-income countries, where there are massive variations in the extent to which households (especially the poor) are protected from catastrophic expenditure. There is therefore increasing interest in WHO and the World Bank developing reliable and comparable indicators of financial protection (16). A major challenge is to move beyond the immediate expenditure on health care, to trace the longer-term implications for household wealth and savings.

Finally, productivity (and efficiency) is perhaps the most challenging measurement area of all, as it seeks to offer a comprehensive framework that links the resources used to the measures of effectiveness described above. The need to develop reliable measures of productivity is obvious, given the policy problems of trying to decide where limited health system financial resources are best spent and of trying to identify inefficient providers. The experience of *The world health report 2000* (7), however, illustrates how difficult this task is at the macro

Box 1. Hospital benchmarking in Finland**Background**

In Finland, in 1997, the National Research and Development Centre for Welfare and Health launched a research and development project (Hospital Benchmarking) to produce benchmarking information on hospital performance and productivity (18). The main aims of the project were:

- to develop a new measure to describe the output of hospitals that was better than traditional measures, such as admissions or outpatient visits; and
- to provide the management of hospitals with benchmarking data for improving and directing activities at hospitals.

Data collection

The project was expanded to cover nearly all publicly delivered specialized health care in Finland and, in 2006, data from the project was integrated into the production of national statistics. Data for the Hospital Benchmarking project are collected annually from hospitals, and they include both inpatient and outpatient care, along with information on diagnoses and procedures. The project produces a wide range of hospital and regional (hospital-, district- and municipality-based) indicators on hospital productivity and costs – by specialty, inpatient wards and diagnosis-related groups. By using uniform personal identity codes, the different episodes of care of the same patient can be linked together.

Uses of data

The data allow regional measurement of productivity and costs, which indicate, for example, how much the costs of a hospital district or a municipality deviate from the national average and how much of this deviation depends on the inefficient delivery of services and the use of services per person.

Hospital Benchmarking data have been used increasingly for appraising and directing hospital activities. Results from the Hospital Benchmarking project indicate that productivity of hospitals decreased somewhat during the years 2001–2005 and that there are significant differences in productivity between hospitals (19).

level. And the accounting challenges of identifying resources consumed become progressively more acute as one moves to finer levels of detail, such as the meso level (provider organizations, for example), the clinical department, the practitioner, or – most challenging of all – the individual patient or person (17). Box 1 gives details of the Finnish experience with producing benchmarking data to use for productivity improvement.

Methodological issues about performance measurement

The diverse uses of health system performance measures necessitate a wide variety of measurement methods, indicators, analytical techniques and approaches to presentation. Also, different methods of data collection – such as national surveys, patient surveys, administrative databases and routinely collected clinical information – are needed to assemble these diverse

Box 2. OECD Health Care Quality Indicators Project**Background**

Since its beginning, in 2001, the OECD Health Care Quality Indicators Project has aimed to track the quality of health care in a number of countries, to assess the quality of international health care. This is done by developing a set of indicators based on comparable data that can be used to investigate quality differences in health care among countries.

Indicators

The five areas in which indicators are being collected are:

1. patient safety
2. quality of mental health care
3. quality of health promotion, illness prevention and primary care
4. quality of diabetes care
5. quality of cardiac care.

The collection of indicators follows a twofold process. Initially, data will be gathered from a limited set of new indicators prepared by teams of internationally renowned experts in each of the five areas. Then country experts in all five areas will conduct work that will provide the basis for improving quality data systems across countries.

Source: Health Care Quality Indicators Project (20).

types of information. The area of performance under scrutiny will determine the most appropriate data collection technique. For example, when measuring responsiveness, household or individual surveys are likely to be the best sources of patient experiences and perspectives, whereas when looking at specific clinical outcomes, clinical registries may be a more informative and cost-effective source of information. In practice, although performance measurement efforts have progressed over recent years, many health systems still rely on readily available data as a basis for performance measurement.

The first requirement in any performance measurement system is to develop a robust conceptual framework within which performance measures can be developed. This should ensure that all major areas of health system performance are covered by the measurement system, that priorities for new developments can be identified and that collection and analysis efforts are not misdirected or duplicated. In short, the eventual requirement is to develop an optimal portfolio of performance-measurement instruments. An example of such a framework is the Organisation for Economic Co-operation and Development (OECD) Health Care Quality Indicators Project, which seeks to assemble a suite of performance indicators that are common to a large number of national performance measurement schemes (Box 2).

Detailed issues about methodology arise when

Table 4. Characteristics of good performance indicators

Stages	Characteristics of indicators
Development of indicators	<i>Face/content validity</i> : the extent to which the indicator accurately measures what it purports to measure
	<i>Reproducibility</i> : the extent to which the indicator would be the same if the method by which it was produced was repeated
Application of indicators	<i>Acceptability</i> : the extent to which the indicator is acceptable to those being assessed and those undertaking the assessment
	<i>Feasibility</i> : the extent to which valid, reliable and consistent data are available for collection
	<i>Reliability</i> : the extent to which there is minimal measurement error or the extent to which findings are reproducible should they be collected again by another organization
	<i>Sensitivity to change</i> : the extent to which the indicator has the ability to detect changes in the unit of measurement
	<i>Predictive validity</i> : the extent to which the indicator has the ability to accurately predict

Source: Adapted from Campbell et al. (21).

considering the design of individual indicators. An important consideration is the level at which to present performance data. Possibilities include the macro level (such as national life expectancy), the meso level (such as post-operative mortality rates in hospitals) and the micro level (such as health outcomes achieved by individual practitioners). Table 4 summarizes some of the characteristics of good indicators. The intention is to develop performance measures that exhibit the characteristics of acceptability, feasibility, reliability, sensitivity to change and validity.

The following sections look more closely at the methodological considerations that need to be taken into account when selecting which indicators to use and how to use and interpret them.

Attribution and causality

Fundamental questions that arise when seeking to interpret many performance data are: what has caused the observed performance and to which practitioners, organizations or agencies should variations in performance be attributed? Hauck, Rice & Smith (22) show that there are immense differences in the extent to which the health system influences performance measures, ranging from a very large effect on responsiveness measures (such as waiting time) to a

Box 3. Key considerations when addressing causality and attribution bias

Users of performance measures should consider the following recommendations when addressing causality and attribution bias.

Reports of research that investigates a possible causal and attributable link between the agents being assessed and the quality outcome proposed should be evaluated with particular attention to:

- the study methodology;
- its controls for confounding variables; and
- the generalizability of the study sample.

Prospective analyses to identify critical pathways involved in the achievement of desired and undesired processes and outcomes of care should be undertaken. These analyses should try to:

- identify possible confounders; and
- identify the extent to which agents under assessment are or can be clustered into homogeneous groupings.

In new performance measurement initiatives, sources of random and systematic error in measurement and sampling should be carefully considered when developing the design. Procedures of data collection that maximize the reliability and accuracy of data (both primary and secondary) used for quality assessment should be institutionalized.

Risk-adjustment techniques should be employed when evaluating the relationship between agents under assessment and the quality indicators. Hierarchical models should be used to account for the clustering of data within different levels of the health system under analysis. The use of statistical methods, such as propensity scores or instrumental variables, should be considered.

Causality and attribution bias cannot be completely eliminated, even when utilizing the best available statistical methods. Unintended effects from biases in assessment of performance should be monitored carefully, especially when reimbursement or other incentives are linked to the measures.

Source: Adapted from Terris & Aron (23).

small effect on population mortality, which is heavily influenced by factors outside the health system. Such variations should be considered when holding providers and other stakeholders to account. To guide policy, improve service delivery and ensure accountability, it is critical that the causality of observed measures is attributed to the correct source(s). When using statistical methods to evaluate causal relationships and guide policy, researchers and policy-makers should be careful to control properly for measurement and attribution bias (23). Box 3 gives key considerations that users of performance measures need to take into account when addressing causality and attribution bias.

Risk adjustment is an approach widely used to address the problem of attribution. It adjusts outcome data according to differences in resources, case mix and environmental factors, thereby seeking to enhance comparability (Box 4). In health care, in particular, variations in patient outcomes will have much to do

Box 4. Statistical considerations when performing risk adjustment

Risk adjustment often involves using statistical modelling applied to large databases with information from many providers. The techniques produce weighting schemes for assessing patient risk. The statistical models can then be used to estimate the *expected* outcome for a provider, given its mix of patients or populations. Its actual outcome is then compared with this benchmark. The following should be considered when performing risk adjustment.

- Optimal risk-adjustment models result from a multidisciplinary effort that involves the interaction of clinicians with statisticians, as well as with experts in information systems and data production.
- Different practice patterns, patient characteristics and data specifications may limit the transferability of models across different countries. Before applying a model developed in another setting, clinicians and methodologists should examine its clinical validity and statistical performance.
- Decision-makers should be wary when drawing conclusions about the performance of risk-adjustment models from statistical summary measures (such as coefficient of determination, R^2 , values), as these measures may not capture the model's predictive ability for different patient subgroups.
- In cases where it is believed that patient characteristics may also influence differences in the treatment patients receive, it may be more appropriate to apply risk stratification instead of (or alongside) risk adjustment.

Source: Adapted from Iezzoni (24).

with variations in patient attributes, such as age or socioeconomic class, and any co-morbidities. Similar considerations apply when comparing measures of population health. In such cases, it is essential to employ methods of risk adjustment when using indicators and comparing agents. A key question then is: for what is the agent under scrutiny accountable? In the short run, for example, a health system has to deal with the epidemiological patterns and risky behaviour it inherits. This implies a major need for risk adjustment when comparing it with other health systems. In the longer run, one might expect the health system to be accountable for improving epidemiological patterns and health-related behaviour. The need for risk adjustment then becomes less critical, as the health system is responsible for many of the underlying causes of the measured outcomes.

Since early efforts with diagnosis-related groups in the United States, the methods of risk adjustment have been steadily refined over a period of 40 years, particularly in adjusting for outcomes for specific diseases or health care treatments. A key issue remains the quality (especially the completeness) of the data on which risk adjustment is undertaken, especially the presence of co-morbidities or other complications. Recording these data depends (ultimately) on the

practitioners whose performance is being assessed, so there is an ever-present threat to the integrity of the data if the incentives associated with performance comparison are too stark. Also, most risk-adjustment efforts are still works in progress, and there is often a need for careful qualitative clinical commentary on any risk-adjusted data, as there are often technical limitations to any scheme. Risk adjustment, however, is almost always essential if performance measurement is to secure credibility with practitioners, so it is important that efforts to improve on current methodologies are sustained.

A specific issue in the interpretation of many performance data is random variation, which by definition emerges with no systematic pattern and is always present in quantitative data. Statistical methods become central to determining whether an observed variation in performance has arisen by chance, rather than from variations in the performance of agents within the health system. As a matter of routine, confidence intervals should be presented alongside performance indicators. In the health care area, a challenge for such methods is to identify genuine outliers in a consistent and timely fashion, without signalling an excessive number of *false positives*. This is crucial when undertaking surveillance of individual practitioners or teams. In dealing with this situation, one must ask: when does a deviation from expected outcomes become a cause for concern and when should a regulator intervene? Statistical methods of squeezing maximum information from time series of data are now reaching an advanced stage of refinement and offer great scope for more focused intervention (25).

Composite measures

Health systems are complex entities with multiple aspects, making performance very difficult to summarize, especially through a single measure. Yet, when separate performance measures are provided for the many different aspects of the health system under observation – such as efficiency, equity, responsiveness, quality, outcomes and access – the amount of information provided can be overwhelming. Such information overload makes it difficult for the users of performance information to make any sense of the data. In response to these problems, the use of composite indicators has become increasingly popular. Composite indicators combine separate performance indicators into a single index or measure and are often used to rank or compare the performance of different practitioners, organizations or systems, by providing a bigger picture and offering a more rounded view of performance (26).

However, if composite indicators are not carefully designed, they may be misleading and could lead to serious failings if used for health system policy-making or planning (27). One of the main challenges in creating composite indicators is deciding which measures to

Table 5. Advantages and disadvantages of composite indicators

Advantages	Disadvantages
<p>Offer a broad assessment of system performance</p> <p>Place system performance at the centre of the policy arena</p> <p>Enable judgement and cross-national comparison of health system efficiency</p> <p>Offer policy-makers at all levels the freedom to concentrate on areas where improvements are most readily secured, in contrast to piecemeal performance indicators</p> <p>Clearly indicate which systems represent the best overall performance and improvement efforts</p> <p>Can stimulate better data collection and analytic efforts across health systems and nations</p>	<p>May disguise failings in specific parts of the health care system</p> <p>Make it difficult to determine where poor performance is occurring and, consequently, may make policy and planning more difficult and less effective</p> <p>Often can lead to double counting, because of high positive correlation</p> <p>May use feeble data when seeking to cover many areas, which may make the methodological soundness of the entire indicator questionable</p> <p>May make individual measures used contentious and hidden, due to aggregation of the data</p> <p>May ignore aspects of performance that are difficult to measure, leading to adverse behavioural effects</p> <p>May only reflect certain preferences when inadequately developed methods for applying weights to composite indicators are used</p>

Source: Adapted from Smith (27).

include in the indicator and with what weights. As composite indicators aim to offer a comprehensive performance assessment, they should include all important aspects of performance, even if they are difficult to measure. In practice, however, there is often little choice of data, and questionable sources may be used for some components of the indicator. Considerable ingenuity may therefore be needed to develop adequate proxy indicators (26, 27).

Fundamental to composite indicators is the choice of weights (or importance) to be attached to the component measures. All the evidence suggests that there exist great variations in the importance different people attach to different aspects of performance, so the specification of a single set of weights is fundamentally a political action. This indicates that the choice of weights requires political legitimacy on the part of the decision-maker. Analysis can therefore inform, but should not determine, the choice of weights. There exists a body of economic methodology for inferring weights, which includes methods for calculating willingness to pay, for eliciting patient's preferences from rankings of alternative scenarios, and for directing making choices in experiments. These economic methods, however, have not been widely applied to the construction of composite indicators of health system performance (27).

Besides capturing effectiveness, a primary benefit of composite indicators is that they allow the construction of measures of the overall productivity (or cost-effectiveness) of a health system. In particular, a

composite measure of health system attainment can be assessed alongside expenditure without the need to assign an expenditure to specific health system activities. This was a principle underlying *The world health report 2000* (7). However, the response to that report emphasized that many aspects of constructing composite attainment and productivity indicators are disputable. Table 5 takes a closer look at the advantages and disadvantages of using composite indicators for health performance assessment.

Using performance measurement: key policy levers

Rapid advances in technology and analytical methodology, coupled with changing public and professional attitudes, have made the use of large-scale information systems for performance assessment and improvement increasingly feasible (4). Experiences with realizing the potential of new data resources to improve system performance, however, have so far shown inconsistent results, and no consensus exists yet on the best way to proceed. This section looks at some of the experiences in using data for performance improvement and at the lessons learned to date.

Information systems

Many of the earliest efforts to use performance data concentrated on collecting and organizing existing administrative information and disseminating it for management applications. These early efforts focused mainly on cost containment and resource allocation. Examples include the development of diagnosis-related

Box 5. The Nordic collaboration**Background**

A Nordic Council of Ministers working group, consisting of 3–4 representatives from each of the Nordic countries (Denmark, Finland, Greenland, Iceland, Norway and Sweden), was established in 2000. Its overall aim is to facilitate collaboration between the Nordic countries through the development of quality indicators and the creation of a foundation for evaluations that should benefit the public, health care professionals and health managers.

Indicators

Six subgroups work on selecting generic and disease-specific indicators and indicators within the areas of patient safety, psychiatry, primary health care, acute somatic care, public health and preventive health care, and patient-experienced health care. So far, the joint quality indicators selected for the Nordic countries fall under the following categories:

1. general and disease-specific indicators (mortality and survival rates for common illnesses);
2. health promotion and ill health prevention;
3. mental health;
4. primary care;
5. patient safety; and
6. the patient experience.

Source: National Board of Health and Welfare (30).

groups to compare hospital costs in the United States and the release of a suite of performance indicators in England to help managers understand how their local health systems compared with the rest of the country. Although (from a managerial perspective) such methods are valuable in better exploiting existing data sources, little attention was given to the use of this information for evaluating external accountability or clinical treatment (28).

Later developments, such as the establishment of the Canadian Institute for Health Information in 1994 and the Nordic collaboration in 2000 (Box 5), used large databases of performance measurement in more creative ways to assist with evidence-based decision-making in health planning and with accountability. Initially, performance data were used mostly by federal and provincial institutions. Reports and summary statistics, however, have increasingly been made available to the public – for example, in the form of the Statistics Canada annual reports. The Canadian Institute for Health Information also focused on analysing the data collected to produce reliable summary indicators, to better understand why trends or patterns emerge and, thus, to best guide policy (29).

Recent technological developments have increased the ability to store a greater volume of information with a greater level of detail, distribute it more widely and

flexibly, and update it more quickly. In the future, the development of the electronic health record – containing all the information on a patient's health history – offers vast potential for capturing performance in many areas. Many challenges, however, need to be addressed if this potential is to be transformed into reality. First, due to the sheer amount of data and the speed at which it can be processed, auditing its accuracy is becoming increasingly important and challenging; the possibility of error carries with it severe implications, if increasing reliance is to be placed on performance data. Second, the constant development of technology calls for continual investment in (and maintenance of) the information infrastructure and entails the need to ensure that the increasing number of information systems are mutually compatible, if their full value is to be exploited. Third, coordination is crucial to ensuring that the information collected is comparable across institutions and settings. Finally, the storage and use of so much information raises ethical concerns about patient privacy (31).

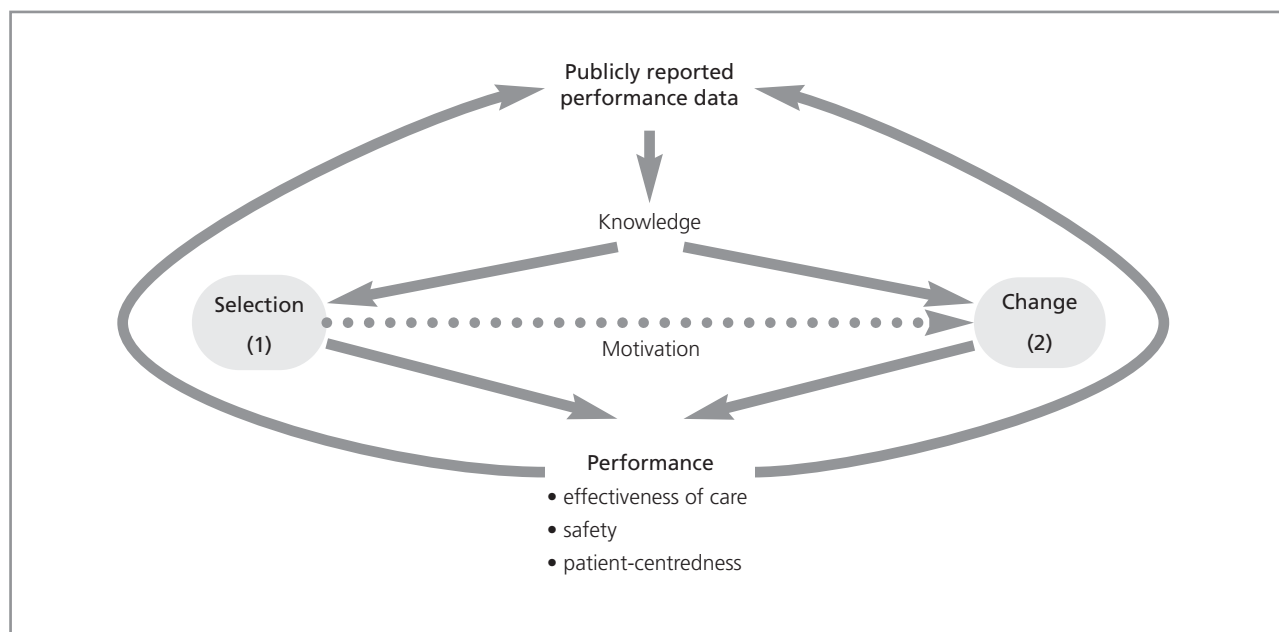
Public reporting

The placement of information in the public domain, to inform the public and other stakeholders about purchaser and provider performance, is growing. This information often takes the form of *report cards* or *provider profiles* that summarizes measures, such as waiting times, patient satisfaction ratings and mortality rates, across providers. Two broad objectives lie behind the public disclosure of information: first, to stimulate quality improvement and, second, to enhance the more general accountability of health system organizations and practitioners to the public who fund and use them. Public reporting can improve quality through two pathways, as illustrated in Fig. 2: (1) a selection pathway, whereby consumers become better informed and select providers of higher quality; (2) a change pathway, whereby information helps providers to identify the areas of underperformance, thus acting as a stimulus for improvement (32).

Both the United Kingdom and the United States have experimented extensively with the use of public disclosure of performance information. The United States has issued report cards for more than 20 years, with its first significant effort led by the federal government agency that administers the Medicare insurance programme. This initiative sought to inform consumer choice and stimulate provider improvement. Following complaints about the validity of the rankings, it was rapidly withdrawn. However, it has since prompted the development of many other performance reports produced by state and federal governments, employers, consumer advocate groups, the media, private enterprise and business purchasers.

There is considerable evidence that publication of provider performance measures leads to improved

Fig. 2. Pathways for improving performance through publicly reported data



Source: Adapted from Berwick, James & Coye (32).

performance (33). Although the immediate purpose of publishing provider performance measures has often been to facilitate and inform patient choice, there is little evidence that patients make direct use of report cards. However, through their effect on the reputation of providers, report cards do appear to promote performance improvements in providers. Apart from their effect on performance, there are growing public demands to make important outcome information public and, in this respect, report cards can assist regulation and enhance accountability.

Starting in 1992, in the United States, two states (New York and Pennsylvania) have experimented with public reporting of post-operative mortality rates for coronary artery bypass graft surgery. These rates are risk adjusted and published at the level of both the hospital and the individual surgeon. The associated confidence intervals are also reported, and a number of empirical analyses have examined the effects of these celebrated report-card initiatives. There is no doubt that the schemes have been associated with a marked improvement in risk-adjusted mortality in the two states (34). However, there is a debate about whether these results necessarily imply that the schemes have been beneficial, and a number of adverse outcomes have also been reported, as follows (35,36).

- The coronary artery bypass graft surgery report cards led to increased selection by New York and Pennsylvania providers, who were more inclined to avoid sicker patients (who might benefit from treatment) and to treat increased numbers of

healthier patients (for whom the benefits of treatment are more contested).

- The initiative has led to increased Medicare expenditures with only a small improvement in population health.
- Practitioners were concerned about the absence of quality indicators other than mortality, about inadequate risk adjustment and about the unreliability of data.

In England, all National Health Service health care organizations are issued an annual performance rating – a report-card rating them from zero to three stars, based on about 40 performance indicators. These ratings were strongly promoted by the national government and received much media and public attention. Poor performance has put executives' jobs at risk, and the initiative has had a strong effect on reported aspects of health care, such as waiting times. However, it has also induced some unintended behavioural consequences, such as a lack of attention to some aspects of clinical quality, which have not been reported. In contrast to the English case, Scotland published a range of important clinical outcome data in the 1990s without any associated publicity. Many governors, clinicians and managers were unaware of the initiative and few incentives were attached to the reports. As a result these indicators had very little impact on the behaviour of practitioners or organizations (37). This experience highlights the need to associate an incentive (which might be financial or reputation- or market-based) with a public-reporting scheme.

Box 6. National quality indicators in Norway**Background**

Norway started to use national quality indicators for specialized health care services in 2003. By 2006, data for 21 indicators were registered (11 for somatic care and 10 for psychiatric care) and, in addition to the indicators, patient-experience surveys were also included. Data reporting is compulsory, and data is published on the Free Hospital Choice Norway web site (38), along with other initiatives and information on the waiting times for different treatments. Data are presented at the hospital level along with data on national averages and developments over time.

Aims

Some important aims of data collection are:

1. to create a base level of quality and generate incentives for health care personnel to improve quality;
2. to identify a base level of quality for management;
3. to support prioritization of health care services by political and administrative entities;
4. to provide the public with information and create transparency in health care services; and
5. to provide users with information to make decisions.

Norway offers another example of public disclosure of performance information. Box 6 discusses the use of national quality indicators in Norway.

Publicly reported information has had a limited direct effect so far on patients and professionals, probably because it is necessarily aggregated and because the indicators reported are limited and inconsistent (39). However, there is increasing evidence that health care organizations do take notice of these data, which have an important effect on their reputations, and that publication of performance information has led to concrete performance improvements (34,40). Notwithstanding doubts about its effectiveness in promoting system improvement, the publication of performance information also serves an important accountability role. There is therefore no doubt that increased public reporting of outcomes of care is an irreversible trend in most health systems. However, it can lead to adverse outcomes, if not implemented and monitored with care.

Experience to date suggests the following points should be taken into account when implementing public disclosure of data.

- Careful consideration should be given to the purpose of the disclosure and to the type of information the different health system stakeholders want and are able to use.
- Careful consideration should be given to the effect that public disclosure of information may have on

quality of care. Where appropriate, public disclosure of information should be integrated with other quality improvement strategies (41).

- To enhance their credibility and usefulness, public performance reports should be created in collaboration with physicians and other legitimate interest groups (35, 41).
- When reporting data, careful risk adjustment should be implemented to offer accurate comparisons between providers and to ensure that the legitimacy of the comparisons is accepted by professionals (24, 41). Detailed information on the risk-adjustment strategies used should be made available alongside the reported information for public scrutiny.

Incentives

There is no doubt that clinicians and other actors in the health system generally respond as expected to financial incentives (42). The incorporation of performance measurement into financial-incentive regimes therefore potentially offers a promising avenue for future policy, and a number of experiments that attach financial rewards to reported performance are now under way.

Historically, the use of indirect financial incentives in health care has been proffered through systems of accreditation that offer rewards in the form of access to markets or extra payments, once specified structures of care are put in place. Germany has an accreditation system of this sort at the regional level, where specific quality indicators are used for accreditation (43). Accreditation is, however, a very blunt incentive instrument. Policy is now shifting towards much more direct and focused incentives. In particular, the United States has been experimenting with financial incentives in different contexts, such as the *rewarding results* experiment, which uses incentives to improve quality (44). However, these have so far been small-scale experiments, and the results have been difficult to assess with any confidence.

Many issues need to be considered when designing performance-incentive schemes, including which aspects of performance to target, how to measure attainment, how to set targets, whether to offer incentives at the individual or group level, how strong to make the link between achievement and reward, and how much money to attach to an incentive. Also, evaluating such schemes is essential, but challenging. In most instances, a controlled experiment is not feasible, as it is often not feasible to establish a convincing *do-nothing* baseline with which to compare the policy under scrutiny. Moreover, constant monitoring is needed to ensure that unintended responses to incentives (such as cream-skimming or other unwanted behavioural responses) are not occurring, that the incentive scheme does not jeopardize the reliability of the performance data on

Box 7. The contract for general practitioners, United Kingdom**Framework**

In April 2004, a new general-practitioner contract took effect in the United Kingdom National Health Service. This new contract more closely linked general practitioners with quality targets for both clinical and organizational activities through the Quality and Outcomes Framework programme. The programme rewards general practitioners for meeting targets in targeted areas, measured by about 150 indicators. Each indicator has a number of points allocated to it, varying according to the amount and difficulty of work required to successfully meet these criteria. A maximum of 1050 points can be earned, and up to 20% of general practitioner income is at risk under the scheme.

Targeted areas

Indicators upon which points are allocated are measured for the following main categories (some smaller categories are omitted):

- clinical areas (76 indicators (focused on medical records, diagnosis, and initial and ongoing clinical management) and 550 points): such as coronary heart disease, stroke and transient ischaemic attack, hypertension, hyperthyroidism, diabetes, mental health, chronic obstructive pulmonary disease, asthma, epilepsy and cancer;
- organizational areas (56 indicators and 184 points): such as records and information about patients, communication with patients, education and training, practice management and medicine management;
- patient experience (4 indicators and 50 points): such as appointment length and consulting with patients about other issues; and
- additional services (10 indicators and 36 points): such as cervical screening, child health surveillance, maternity services and contraceptive services.

No risk adjustment is undertaken. Instead, practices may exclude certain patients from performance measurement, if the required intervention is clinically inappropriate or if the patient refuses to comply.

Findings to date

- In preparation for the 2004 programme, general practitioners in the United Kingdom employed more nurses and administrative staff, established chronic-disease clinics and increased the use of electronic medical records (46). Also, general practitioners are increasingly delegating tasks to other members of clinical staff. For example a nurse may be asked to specialize in diabetes care (47).
- Although the Quality and Outcomes Framework programme was voluntary, in its first year of implementation almost all United Kingdom practices chose the programme, with the median practice scoring 95.5% of the possible points available. In the clinical areas, the median score was 96.7% (46). The achievements of years two and three of the contract have been similarly high (48).
- Interviews with general practitioners suggested they were concerned about the programme's focus on biomedical targets, which may lead to a reduced focus on other important aspects of care and may interfere with their ability to treat the patient as a whole person (47).
- There is little evidence of manipulation of the prevalence data on which performance is based. However, some practices do appear to be making excessive use of exception reporting (49).
- Although there is some evidence that the Quality and Outcomes Framework programme has improved patient care, quality was already improving rapidly in primary care and the specific effect of the programme seems to have been small (50, 51).

Source: Adapted from Lester & Roland (52).

which it relies, and that it does not compromise unrewarded aspects of performance.

The United Kingdom is experimenting with an ambitious financial-reward system for general practitioners, introduced in April 2004, under which about 20% of earnings are directly related to their performance across about 150 quality indicators (45) (Box 7). So far, it has not been possible to attribute any major improvements in general-practitioner performance, or other system improvements, to this bold (and very expensive) experiment. More generally, while performance-based incentive schemes do appear to offer immense potential for system improvement, there is a clear need for more careful research to identify the best mechanisms for harnessing their potential.

Targets

Health system targets are a specific type of performance measurement and incentive scheme and are a quantitative expression of an objective to be met in the future. Targets have been brought to health policy from the field of business, the main idea being that when goals are explicitly defined as targets, more organized and efficient efforts will be made to meet them. Targets are expected to be SMART: specific, measurable, accurate, realistic and time bound (53). If well designed, targets can help organizations and practitioners focus on a manageable number of achievable goals, which thereby lead to system improvements. The governments of many countries – including European Region Member States (most notably, the United Kingdom), Australia,

New Zealand and the United States – have experimented with targets in health care.

However, evidence on the success of using health system targets is limited (54). They have traditionally been used extensively in public health, but reports of measurable success are rare. The English experience with the 1992 Health of the Nation strategy is typical. The strategy was based on the WHO health for all initiative and set a series of ambitious public health targets. However, a careful independent evaluation in 1998 concluded that its “impact on policy documents peaked as early as 1993; and, by 1997, its impact on local policy-making was negligible” (55). Hunter summarized its failings under six broad headings (56).

1. There appeared to be a lack of leadership in the national government.
2. The policy failed to address the underlying social and structural determinants of health.
3. The targets were not always credible and were not formulated at a local level.
4. The strategy was poorly communicated beyond the health system.
5. The strategy was not sustained.
6. Partnerships between agencies were not encouraged.

In the past decade, targets have been an especially strong feature of English health care policy. Starting in 1998, the Treasury issued strategic targets, called Public Service Agreements, to all government departments, including the health ministry (57). Public Service Agreements were focused primarily on outcomes, such as the improvement of mortality rates, reductions in smoking and obesity, and reductions in waiting times. The health ministry used the star rating report cards, described above, as a key instrument to achieve these objectives. In contrast to most national target systems, this proved notably effective in securing some of the targeted objectives in health care (58). This success can be attributed to the following.

- The targets were precise, short-term objectives, rather than long-term and general.
- Targets were based at the local level, rather than the national level.
- Professionals were engaged in the design and implementation of some of the targets. While this ran the risk of leading to so-called capture by professional interests, it also served to increase the awareness of objectives.
- Organizations were given increased financing, information and managerial capacity to respond to challenging targets.
- Concrete incentives were attached to the targets.

Box 8. Risks associated with increased reliance on targets

- Untargeted aspects of the health system may be neglected.
- Managers and practitioners may concentrate on short-term targets directly in their control at the expense of targets that address long-term or less controllable objectives.
- The complexity of the target system requires a large implementation capacity and may be influenced by professional interests.
- Excessively aggressive targets may undermine the reliability of the data on which they are based.
- Excessively aggressive targets may induce undesirable behavioural responses.
- Targets may encourage a narrow, mercenary attitude, rather than encouraging altruistic professionalism.

Source: Smith (59).

However, this success in health care was not replicated in the area of public health, almost certainly because managers felt health care targets were much more amenable to health system intervention.

While targets provide a straightforward way of highlighting key objectives and can be very successful if designed and implemented correctly, there are notable risks associated with their use (59). Box 8 identifies some of the risks associated with increased reliance on targets. The conclusions from this experience are that, while performance targets offer some latitude for focusing system attention on specific areas of endeavour, they are unlikely to secure performance improvements, unless implemented carefully alongside other improvement initiatives, such as more general inspection and regulation.

Professional improvement

Most of the uses of performance measurement described above have been concerned with providing some means of external assessment and scrutiny of the health system, as a mechanism for prompting improved performance. Yet, another important use of performance measurement can be to provide feedback for clinicians on their performance relative to their peers. Databases that serve this purpose exist in many countries. For example, in Sweden they take the form of *quality registers*, where individual-based data on patient characteristics, diagnoses, treatments, experiences and outcomes are all collected voluntarily on the part of the health care providers and shared with other members of the register. The explicit aim of the quality registers is to facilitate the improvement of quality in clinical work through continuous learning and development (60) (Box 9). Indeed there is a strong argument that performance measurement should become an inherent part of a clinician’s lifelong learning. This suggests the need for a

Box 9. Sweden's quality registers

The development of national quality registers in Sweden has been a major effort in promoting performance improvement. Sweden has about 50 active quality registers, with the first one dating back to 1979.

The aim of a national quality register is to encourage good medical practice through the comparison and evaluation of outcome and quality information over time and between providers.

A variety of organizational patterns are used, but each is clinically led and typically maintained by a group (usually located in one of the Swedish university hospitals) that collects, assembles, analyses and distributes the data to its members. Several meetings might be organized each year to discuss this material. The participation of clinicians in a registry group is voluntary, and in most cases registers develop gradually.

When a register is developed, the quality indicators and reporting tools are established on the basis of consensus within the medical specialty and are often refined from year to year. Information on departments is anonymous. However, most well-established registers do present department data publicly. The quality registers provide clinicians with essential information with which to compare performance and facilitate discussion on improvement. Increasingly, data from quality registers have also been used to support decision-making.

Source: Rehnqvist (60).

prominent role for performance measurement principles in early clinical training.

Whether information for professional improvement should be kept anonymous or be made available to the public is widely debated. Evidence suggests that, to be effective, such performance measurement schemes need to be designed and owned by the professionals who use them (61). It is argued that the most constructive systems are those that encourage positive and cooperative behaviour among clinicians and avoid public threats to their professional or commercial standing, which may encourage defensive behaviour that could lead to cream skimming or other unwanted behavioural responses. Indicators used for professional improvement should therefore:

- reflect meaningful aspects of clinical practice with a strong scientific underpinning;
- ensure risk adjustment of indicators;
- allow exclusion of certain patients, such as those who refuse to comply with treatment;
- facilitate interpretability;
- represent services under a provider's control;
- ensure high accuracy; and
- minimize cost and burden.

Also, as well as measuring the outcomes of care, it is

important to seek to measure the extent of inappropriate care (overuse or underuse of treatments).

The requirements of a successful professional-improvement performance-measurement system may therefore come into conflict with the requirements of information systems designed to promote accountability and patient choice. This is not to say that the tension between these different needs and demands cannot be resolved. Experiences from Sweden and elsewhere, such as Denmark and the Netherlands, suggest that public and professional needs can be reconciled – for example, some quality registers do publish outcomes on individual practitioners (62). In any case, patients will in all likelihood increasingly demand that more performance data be made available. The challenge for the professions is to ensure that this trend is harnessed to good results, rather than leading to defensive professional behaviour. One solution lies in the careful development of acceptable, statistical, risk-adjustment schemes and in careful presentation of statistical data, so that the public and media are better equipped to understand and interpret the information that is made available to them.

Summary and conclusions

The ultimate goal of any performance-measurement instrument is to promote the achievement of health system objectives. Thus, its effectiveness should be evaluated not in relation to statistical properties, such as accuracy and validity, but should be evaluated more broadly in relation to the extent to which it promotes or compromises these objectives. Effective performance measurement alone is not enough to ensure effective performance management. The functions of analysis and interpretation of performance data are also crucial. Also, performance measurement is only one (albeit very important) instrument for securing system improvement. To maximize its effect, performance measurement needs to be aligned with other aspects of system design, such as financing, market structure, accountability arrangements and regulation. Finally, a great deal of attention needs to be paid to the political context within which any performance-measurement scheme is implemented. Without careful attention to these broader health system considerations, the best performance-measurement system will be ineffective.

Governments have a major stewardship role to play in harnessing the full potential of performance measurement for improving the health system. *The world health report 2000* (7) defined stewardship as “defining the vision and direction of health policy, exerting influence through regulation and advocacy, and collecting and using information”. The present document has sought to outline how performance measurement can help governments fulfil each of these roles. It has argued that performance measurement offers health systems major opportunities to secure performance improvement and that no health system

can be adequately steered without good performance information and intelligence. The overarching role of performance measurement is to enhance the decisions made by actors throughout the health system.

Performance information can help governments directly in formulating and evaluating policy and in undertaking regulation. The broader stewardship role of governments is, however, to ensure that the necessary flow of information is available, functioning properly and aligned with the design of the health system. Performance measurement is a public good that will not occur naturally. Governments therefore have a fundamental role to ensure that the maximum benefit is secured from performance measurement, whether through law, regulation, coordination or persuasion. Implementation then requires sustained political and professional leadership at the highest level and also assurance that the necessary analytical capacity is available throughout the health system.

Some of the stewardship responsibilities of government in the area of performance measurement are summarized in Box 10.

Given the increasing demand for performance measurement and given the large set of actors and responsibilities, it is important that policy-makers consider what makes performance indicators effective in improving system performance and accountability. Although there is no conclusive answer to this question, experience has suggested that any policy development should embrace the following.

1. A clear conceptual framework and a clear vision of the purpose of the performance-measurement system should be developed and should be aligned with the accountability relationships inherent in the health system.
2. Performance indicators should attempt to measure performance that is directly attributable to an organization or actor, and not to environmental factors (such as patient attributes or socioeconomic factors).
3. Definitions of performance indicators should be clear and consistent and should fit into the conceptual framework chosen.
4. Indicators should aim to measure concepts that are relevant to the needs of specific actors and should not focus merely on measuring what is available or easy to measure.
5. Indicators should aim to be statistically sound and should be presented in a way that is straightforward to interpret, thus reducing the likelihood of manipulation or misinterpretation.
6. Indicators should be presented with full acknowledgement of any data limitations, including uncertainty estimates and lack of

Box 10. Stewardship responsibilities associated with performance measurement

Stewardship responsibilities associated with performance measurement can be summarized under the headings that follow. None of these roles need be undertaken by government itself, but it must be ensured that they all function effectively:

1. **development of a clear conceptual framework and a clear vision of the purpose of the performance measurement system:**
 - alignment with accountability relationships;
 - alignment with other health system mechanisms, such as finance, market structure and information technology;
2. **design of data collection mechanisms:**
 - detailed specification of individual indicators;
 - alignment with international best practice;
3. **information governance:**
 - data audit and quality control;
 - ensuring public trust in information;
 - ensuring well-informed public debate;
4. **development of analytical devices and capacity to help understand the data:**
 - ensuring analysis is undertaken efficiently and effectively;
 - ensuring local decision-makers understand the analysis;
 - commissioning appropriate research on, for example, risk adjustment, uncertainty and data feedback mechanisms;
5. **development of appropriate data aggregation and presentational methods:**
 - ensuring information has appropriate effect on all parties;
 - mandating public release of summary comparative information;
 - ensuring comparability and consistency;
6. **design of incentives to act on performance measures:**
 - monitoring effect of performance information on behaviour;
 - acting to enhance beneficial outcomes and negate any adverse consequences;
7. **proper evaluation of performance-measurement instruments:**
 - ensuring money is spent cost-effectively on information resources;
8. **managing the political process:**
 - developing and monitoring policy options;
 - encouraging healthy political debate;
 - ensuring that specific interest groups do not capture the performance information system.

timeliness. Further exploration of improved processes for handling measurement errors is needed, as such errors may confound true performance differences.

7. More attention should be paid to the presentation of performance data and how this influences their

interpretation by patients, providers and provider organizations.

8. Attention should be given to enhancing the capacity to understand and use information among managers and clinicians. Use of performance data should become an intrinsic part of clinical education and lifelong professional development.
9. Incentives that act on performance measures should be carefully designed. The impact of performance information on behaviour should be carefully monitored, and actions should be taken to enhance beneficial outcomes and to negate any adverse consequences.
10. Policy-makers should pay particular attention to the broader health system, to ensure that performance measurement is aligned with the design of mechanisms, such as finance and market structures, and to recognize the organizational context within which performance data are collected and disseminated.
11. Performance measurement systems should be monitored frequently and evaluated to identify opportunities for improvement and any unintended side-effects.
12. The political aspects of performance measurement should be managed effectively. Among other things, this involves ensuring that specific interest groups do not capture the performance information system and also involves encouraging healthy political debate.

Health systems are still in an early stage of performance measurement, and major steps can still be taken to improve the effectiveness of their measurement systems. Performance measurement, however, offers opportunities for major health system improvements. Advances in technology are likely to increase this potential still further, and the increasing public demands for accountability and information will reinforce current trends. There is therefore a policy-making imperative to consider carefully the role of performance measurement in the health system, to implement initiatives of proven effectiveness, to undertake careful trials of less established mechanisms and to monitor and update performance measurement systems as new knowledge and capacity emerge.

References

1. Loeb JM. The current state of performance measurement in health care. *International Journal for Quality in Health Care*, 2004, 16(Suppl. 1):i5–i9.
2. McIntyre D, Rogers L, Heie E. Overview, history, and objectives of performance measurement. *Health Care Financing Review*, 2001, 22:7–43.
3. Spiegelhalter DJ. Surgical audit: statistical lessons from Nightingale and Codman. *Journal of the Royal Statistical Society*, 1999, 162(1):45–58.
4. Power M. *The audit society: rituals of verification*, 2nd ed. Oxford, Oxford University Press, 1999.
5. Smith PC. Performance measurement in health care: history, challenges and prospects. *Public Money & Management*, 2005, 25(4):213–220.
6. Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).
7. *The world health report 2000: health systems: improving performance*. Geneva, World Health Organization, 2000 (<http://www.who.int/entity/whr/2000/en/index.html>, accessed 13 May 2008).
8. Holland W. *European Community atlas of 'avoidable death'*. Oxford, Oxford University Press, 1988.
9. Nolte E, McKee M. *Does healthcare save lives? Avoidable mortality revisited*. London, The Nuffield Trust, 2004.
10. Fitzpatrick R. Individual health outcomes from health care interventions. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).
11. Donabedian A. Evaluating the quality of medical care. *Milbank Memorial Fund Quarterly*, 1966, 44(3):166–206.
12. Naylor DC, Iron K, Handa K. Measuring health system performance: problems and opportunities in the era of assessment and accountability. In: Smith PC, ed. *Measuring up: improving health system performance in OECD countries*. Paris, Organisation for Economic Co-operation and Development, 2002:13–34.
13. Davies H. *Measuring and reporting the quality of health care: issues and evidence from the international research literature*. Edinburgh, NHS Quality Improvement Scotland, 2005 (<http://www.nhshealthquality.org/nhsqis/files/Davies%20Paper.pdf>, accessed 13 May 2008).
14. Mant J. Process versus outcome indicators in the assessment of quality of health care. *International Journal for Quality in Health Care*, 2001, 13(6):475–480.
15. Valentine N et al. Responsiveness. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).
16. Wagstaff A. Financial protection. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).
17. Street A, Hakkinen U. Health system. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).
18. Linna M. *Benchmarking hospital productivity*. Berlin, Bertelsmann Foundation Health Policy Monitor, 2006 (http://www.hpm.org/en/Surveys/STAKES/07/Benchmarking_hospital_productivity.html, accessed 13 May, 2008).
19. *Sairaaloiden tuottavuuden kehitys 2001–2005 [The development of hospitals' profitability, 2001–2005]*. Helsinki, Finnish National Research and Development Centre for Welfare and Health, 2007 (Tilostotiedote Statistikkimeldande 5/2007; http://www.stakes.fi/tilastot/tilastotiedotteet/2007/Tt05_07.pdf, accessed 13 May 2008).
20. Health Care Quality Indicators Project [web site]. Paris, Organisation for Economic Co-operation and Development (http://www.oecd.org/document/34/0,3343,en_2649_201185_37088930_1_1_1_1,00.html, accessed 13 May 2008).
21. Campbell SM et al. Research methods used in developing and applying quality indicators in primary care. *Quality and Safety in Health Care*, 2002, 11:358–364.
22. Hauck K, Rice N, Smith P. The influence of health care organisations on health system performance. *Journal of Health Services Research and Policy*, 2003, 8(2):68–74.
23. Terris DD, Aron DC. Attribution and causality in health care performance measurement. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).
24. Iezzoni, L. Risk adjustment for performance measurement. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).
25. Grigg O, Spiegelhalter D. Clinical surveillance and patient safety. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).

26. Goddard M, Jacobs R. Using composite indicators to measure performance in health care. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).
27. Smith PC. Developing composite indicators for assessing health system efficiency. In: Smith PC, ed. *Measuring up: improving the performance of health systems in OECD countries*. Paris, Organisation for Economic Co-operation and Development, 2002:295–316.
28. Smith PC. The use of performance indicators in the public sector. *Journal of the Royal Statistical Society*, 1990, 153(1):53–72.
29. Wolfson M, Alvarez R. Towards integrated and coherent health information systems for performance monitoring: the Canadian experience. In: Smith PC, ed. *Measuring up: improving health system performance in OECD countries*. Paris, Organisation for Economic Co-operation and Development, 2002:133–155.
30. *Nationella kvalitetsindikatorer inom hälso- och sjukvården [National Quality Indicators in the health care system]*. Stockholm, National Board of Health and Welfare, 2006 (<http://www.socialstyrelsen.se/NR/rdonlyres/28769591-A924-4AD9-ADD8-352B79875D51/5783/200610713.pdf>, accessed 13 May 2008).
31. Sequist T, Bates D. Information technology and performance measurement. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).
32. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. *Medical Care*, 2003, 41(Suppl. 1):I30–I38.
33. Hibbard JH, Stockard J, Tusler M. Hospital performance reports: impact on quality, market share, and reputation. *Health Affairs*, 2005, 24(4):1150–1160.
34. Shekelle PG. Public reporting of quality information. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).
35. Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care. *New England Journal of Medicine*, 1996, 335(4):251–256.
36. Dranove D et al. Is more information better? The effects of “report cards” on health care providers. *Journal of Political Economy*, 2003, 111(3):555–588.
37. Mannion R, Goddard M. Impact of published clinical outcomes data: case study in NHS hospital trusts. *British Medical Journal*, 2001, 323:260–263.
38. Fritt Sykehusvalg Norge [Free Hospital Choice Norway] [web site]. Oslo, Social- og Helsdirektoratet (<http://www.frittSykehusvalg.no>, accessed 13 May 2008).
39. Marshall MN et al. Public reporting on quality in the United States and the United Kingdom. *Health Affairs*, 2003, 22(3):134–148.
40. Marshall MN et al. The public release of performance data: what do we expect to gain? A review of the evidence. *Journal of the American Medical Association*, 2000, 283(14):1866–1874.
41. Marshall MN et al. Public disclosure of performance data: learning from the US experience. *Quality in Health Care*, 2000, 9:53–57.
42. Dudley RA. Pay-for-performance research: how to learn what clinicians and policy makers need to know. *Journal of the American Medical Association*, 2005, 294(14):1821–1823.
43. Der Gemeinsame Bundesausschuss [web site]. Siegburg, Gemeinsamer Bundesausschuss (<http://www.g-ba.de/>, accessed 13 May 2008).
44. *Ensuring quality providers: a purchaser's toolkit for using incentives*. Washington, National Health Care Purchasing Institute, 2002.
45. Smith PC, York N. Quality incentives: the case of U.K. general practitioners. *Health Affairs*, 2004, 22(3):112–118.
46. Doran T et al. Pay-for-performance programs in family practices in the United Kingdom. *New England Journal of Medicine*, 2006, 355(4):375–384.
47. Roland MO et al. Financial incentives to improve the quality of primary care in the UK: predicting the consequences of change. *Primary Health Care Research and Development*, 2006, 7(1):18–26.
48. Burr T. *NHS pay modernisation: new contracts for general practice services in England*. London, National Audit Office, 2008 (http://www.nao.org.uk/publications/nao_reports/07-08/0708307.pdf, accessed 13 May 2008).
49. Gravelle H, Sutton M, Ma A. *Doctor behaviour under a pay for performance contract: evidence from the Quality and Outcomes Framework*. York, Centre for Health Economics, University of York, 2007 (CHE Research Paper 28; <http://www.york.ac.uk/inst/che/pdf/rp28.pdf>, accessed 13 May 2008).
50. Campbell S et al. Quality of primary care in England with the introduction of pay for performance. *New England Journal of Medicine*, 2007, 357(2):181–190.
51. Hippisley-Cox J, Vinogradova Y, Coupland C. *Time series analysis for selected clinical indicators from the Quality and Outcomes Framework 2001–2006*. Leeds, The Information Centre for Health and Social Care, 2007.

52. Lester H, Roland M. Performance measurement in primary care. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press (in press).
53. van Herten LM, Gunning-Schepers LJ. Targets as a tool in health policy. Part 1: lessons learned. *Health Policy*, 2000, 53(1):1–11.
54. Wismar M et al., eds. *Good governance for implementing health targets*. Maidenhead, Open University Press (in press).
55. *Health of the nation – a policy assessed*. London, Department of Health, 1998.
56. Hunter D. England. In: Marinker M, ed. *Health targets in Europe: polity, progress and promise*. London, BMJ Books, 2002:148–164.
57. Smith PC. Performance budgeting in England: public service agreements. In: Robinson M, ed. *Performance budgeting: linking funding and results*. Houndmills, Palgrave Macmillan, 2007:211–233.
58. Bevan G, Hood C. Have targets improved performance in the English NHS? *British Medical Journal*, 2006, 332:419–422.
59. Smith PC. Targets in the English health system. In: Wismar M et al., eds. *Good governance for implementing health targets*. Maidenhead, Open University Press (in press).
60. Rehnqvist N. Improving accountability in a decentralized system: a Swedish perspective. In: Smith PC, ed. *Measuring up: improving the performance of health systems in OECD countries*. Paris, Organisation for Economic Co-operation and Development, 2002:87–96.
61. Rowan K, Black N. A bottom-up approach to performance indicators through clinician networks. *Health Care UK*, 2000, Spring:42–46.
62. Den Danske Kvalitetsmodels organisation [The Danish Quality Model Organization] [web site]. Århus, Danish Institute for Quality and Accreditation in Healthcare, 2007 (<http://www.kvalitetsinstitut.dk/sw179.asp>, accessed 13 May 2008).

World Health Organization
Regional Office for Europe
Scherfigsvej 8,
DK-2100 Copenhagen Ø,
Denmark
Tel.: +45 39 17 17 17.
Fax: +45 39 17 18 18.
E-mail: postmaster@euro.who.int
Web site: www.euro.who.int



WHO European Ministerial
Conference on Health Systems:
"HEALTH SYSTEMS.
HEALTH AND WEALTH"

Tallinn, Estonia, 25–27 June 2008

This report is one of three background documents prepared for the WHO European Ministerial Conference on Health Systems: "Health Systems, Health and Wealth", held on 25–27 June 2008 in Tallinn, Estonia. Together, these reports demonstrate that:

- ill health is a substantial burden economically and in terms of societal well-being;
- well-run health systems can improve health and well-being, and contribute to wealthier societies, and
- strategies are available to improve health systems' performance.

These are the key themes of the Conference. These detailed syntheses highlight important research findings and their implications, and underline the challenges that they pose for policy-makers. They support the Conference position that cost-effective and appropriate spending on health systems is a good investment that can benefit health, wealth and well-being in their widest senses.

These three background documents together provide the theoretical foundations around which the aims, arguments and rationale for the Conference are oriented. Document 1 gives the background evidence on the cost of ill health and is supported by twin volumes on health as a vital investment in eastern and western Europe. Documents 2 and 3 represent concise synopses of the two comprehensive Conference volumes being coordinated by the European Observatory on Health Systems and Policies. These volumes on health systems, health and wealth and performance involve a range of leading experts and will be made available to delegates in draft for comment. They will be revised in light of feedback before publication at the end of 2008.

Background document #2

Performance measurement for health system improvement: experiences, challenges and prospects

This summary makes the case for performance measurement as key tool for policy-makers endeavouring to improve health systems in the European Region. It highlights the various elements required of a comprehensive health system performance measurement framework; pinpoints how performance measurement can be used in practice; and stresses the role of government stewardship in securing improved performance. It reviews existing evidence and provides examples of the empirical application of performance measures, demonstrating that if governments invest in health they can expect those resources to be used well.