

Conclusions

6.1 | *Conclusions*

PETER C. SMITH, ELIAS MOSSIALOS,
IRENE PAPANICOLAS, SHEILA LEATHERMAN

In the opening chapter we argue that the goals of any performance measurement instrument are twofold: to promote accountability and to improve the performance of the health system. The modern health system is immensely complex, comprising of diverse agents such as insurers, provider organizations, health-care professionals and central and local governments. Measurement of the actions and outcomes of these agents is a necessary condition if the health system is to be held properly to account by citizens and patients. That accountability may be considered a good thing in its own right as it enhances transparency and promotes informed debate about the health system. Furthermore, by providing reassurance that finances are being used effectively, performance measurement can increase government and citizens' willingness to invest additional resources in the health system. In this book the prime focus is how performance measurement and the increased accountability it offers directly promotes the achievement of health system objectives – higher quality and more cost-effective health care and improved population health.

Measurement alone is not sufficient to achieve these objectives. In this book we cite numerous instances of technically satisfactory performance measurement initiatives that have failed to make material impacts on health systems (or indeed have had perversely adverse impacts). For example, there are examples of public performance reporting schemes being ignored; professional improvement efforts becoming moribund; and the use of centrally mandated targets inducing perverse results. To have maximum effect, performance measurement needs to be aligned with other aspects of system design such as financing, market structure, governance arrangements and regulation. Moreover, great 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.

The effectiveness of any performance measurement initiative should be evaluated not only in relation to (often important) statistical properties such as accuracy and validity but also more broadly – by the extent to which it promotes or compromises broader health system objectives. This book has sought to reflect this broader view of performance measurement. Part 2 describes some recent major technical advances in seeking to measure aspects of health system performance. Part 3 examines some of the analytical techniques currently used to gain a greater understanding of the information contained in performance measures, whilst Part 4 examines advances in some particularly challenging areas of the health system. Part 5 seeks to complete the accountability cycle by examining some of the policy initiatives that have been introduced to promote more effective use of performance measures.

In this chapter we draw out the most important lessons for policy-makers. We begin by emphasizing the need for a conceptual framework to inform the development of performance measurement. Such a framework must be in place in order to undertake a systematic choice of performance indicators, as discussed in the next section. We then examine statistical issues that must be addressed satisfactorily if performance measurement is to be effective and the necessity for indicators to be embedded within an appropriate set of incentives. We go on to discuss the intrinsically political nature of performance measurement, as noted by many authors. In the penultimate section we examine government's role in promoting, facilitating and implementing performance measurement. The concluding section summarizes what we consider to be the main priorities for any health system seeking to improve the measurement of performance.

Conceptual framework

We believe that a fundamental requirement for any performance measurement system is the development of a robust conceptual framework within which specific performance measures can be developed, tested and implemented routinely. The framework should ensure that all major domains of health system performance are covered. The chapters in Part 2 offer an oversight of the main categories of measures that are likely to be useful in most systems. They are summarized below.

Align with health system objectives

It is important that the conceptual framework for performance measurement is aligned with other aspects of health system design. Important considerations might be the payment system; market structure; accountability and governance arrangements; IT infrastructure; and regulation. For example, if a DRG payment system is used it may be sensible to ensure that certain performance measures are consistent with DRG codes. This will enable provider performance to be linked directly with expenditure and will facilitate judgments about efficiency.

Integrate with IT and routine data collection

The link between the performance measurement framework and health system IT arrangements is critical. Rapid changes 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 (Power 1999). So far, there has been patchy and largely idiosyncratic experience of realizing the potential of new data sources to improve system performance; with little consensus across countries and disparate health systems in technical development. Yet technology has transformed the capacity to store a greater volume of information at a great level of detail; distribute this widely, rapidly and flexibly; and update it quickly. The development of the electronic health record, containing all information on a patient's health history, offers vast potential for capturing performance in many areas.

Sequist and Bates (Chapter 5.3) show that many challenges need to be addressed if such potential is to be transformed into reality. First, the sheer amount of data and the speed at which they can be processed makes it increasingly important and challenging to audit their accuracy. If increasing reliance is to be placed on performance data then the possibility of error carries severe implications. Second, the constant development of technology calls for continual infrastructure investment and maintenance. There will be a need to ensure that the increasing numbers of information systems are mutually compatible if their full value is to be exploited. Policy-makers should work to ensure

smooth implementation of IT systems that do not disrupt workflow or hinder efficiency in the short term. Third, there is a crucial coordination role in ensuring that information collected is comparable across institutions and settings. Finally, the storage and use of so much information raises ethical concerns about individual privacy. In short, IT strategy and the performance measurement framework should be considered as an integrated system, developed jointly rather than in isolation.

Include high-priority hard-to-measure areas

Part 4 of the book highlights progress in certain hard-to-measure parts of the health system: primary care, chronic care, mental illness and long-term care. We believe these to be especially important priorities – they represent major expenditure commitments in most health systems, although clinical practice (and therefore the outcomes of services) is especially variable. Furthermore, without adequate performance measurement it becomes very difficult to identify what works in these challenging domains. Paradoxically, it is this shortage of evidence that makes these domains such high priorities for future initiatives.

More generally, the conceptual framework is intended to help identify priorities for new developments and to ensure that collection and analysis efforts are neither misdirected nor duplicated. In short, the eventual requirement is to develop an optimal portfolio of performance measurement instruments that fits a health system's existing organizational structure and accountability arrangements and the available levels of resources and analytical capacity. This may seem a demanding requirement. However, the alternative to maintaining the necessary holistic view will be continued fragmentation and underperformance in some parts of the health system

Design for international comparability

One final consideration in the development of a conceptual framework is the increasing need to harmonize national data with international practice and standardize the definitions of indicators that are compared internationally. The OECD HCQI project is assembling a suite of performance indicators that are common to a large number of national performance measurement schemes (Box 6.1.1), leading

Box 6.1.1 OECD HCQI project

Background

Begun in 2001, the OECD HCQI project aims to assess international health-care quality by developing a set of indicators based on comparable data that can be used to investigate quality differences in health care amongst countries.

Indicators

Indicators are being collected in five areas:

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

The collection of indicators is a two-fold process. Firstly, data are gathered from a limited set of new indicators prepared by teams of internationally renowned experts in each of the five fields. Secondly, country experts in all five areas conduct focus work that will provide the basis for improving quality data systems across member countries.

Source: OECD web site (<https://www.oecd.org/health/hcqi>)

to increased potential for international comparison. Such comparison makes an especially strong contribution to national accountability and is one of the most important stimuli for policy reform.

Choosing performance measures

The selection of performance measures is not only critical for sound assessment but also plays a larger role in defining what is considered important at every level of a health system. In Part 2 we have summarized health system objectives under a limited number of headings such as the health conferred on citizens by the health system; responsiveness to citizen preferences; financial protection offered by the health system; and health system productivity. Furthermore, as well

as a concern with the overall attainment in each of these domains, we highlight the importance of distributional (or equity) issues expressed in terms of inequity in health outcomes, in responsiveness and in payment. Table 6.1.1 summarizes these largely universal dimensions of health performance measurement considered in this book and some example indicators in each.

The chapters show that there is variable progress in the development of performance measures and data collection techniques in the different dimensions of health performance. Some areas (e.g. population health) have well-established indicators such as infant mortality and life expectancy (sometimes adjusted for disability). Yet even here there is scope for important further work. With population health measures there is a particular difficulty in estimating the health system's specific contribution to health. Chapter 2.1 highlights the devel-

Table 6.1.1 *Dimensions of health performance measures*

Measurement area	Description of measure	Examples of indicators
Population health	Measures of aggregated data on the health of the population.	Life expectancy Years of life lost Avoidable mortality DALYs
Individual health outcomes	Measures of individual's health status; can be relative to the whole population or amongst groups. Some indicators also apply utility rankings to different health states.	Generic measures: <ul style="list-style-type: none"> • SF-36 • EQ-5D Disease specific measures: <ul style="list-style-type: none"> • Arthritis Impact Measurement Scales • PDQ-39
Clinical quality and appropriateness of care	Measures of the services and care patients receive to achieve desired outcomes. Used to determine if best practice takes place and that these actions are carried out in a technologically sound manner	Outcome measures: <ul style="list-style-type: none"> • health status • specific post-operative readmission and mortality rates Process measures: <ul style="list-style-type: none"> • frequency of blood pressure measurement

Table 6.1.1 *cont'd*

Responsiveness of health system	Measures of the way individuals are treated and environment in which they are treated during health system interactions. Responsiveness is 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 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 productivity of the health-care system, health-care organizations and individual practitioners.	Labour productivity Cost-effectiveness measures (i.e. for interventions) Technical efficiency (measures of output/ input) Allocative efficiency (i.e. measured by willingness to pay)

opment of more recent instruments such as the concept of *avoidable* mortality (Holland 1988; Nolte & McKee 2004).

In the health-care domain Chapter 2.2 notes increasing interest in measures of improvements in patient health status, often in the form of PROMs, derived from simple surveys of subjective health status administered directly to patients. There is now a plethora of measurement

instruments ranging from detailed condition-specific questionnaires to broad-brush generic measures of patient outcome. For performance measurement purposes one central policy challenge is to identify the most appropriate choice of instrument. For example, in England the government recently mandated the use of the EQ-5D generic PROM instrument for all NHS patients undergoing four common procedures: hip replacement, knee replacement, hernia repair and varicose vein surgery. This experiment will assess the feasibility and costs of such routine use and test whether the resistance to PROMs among some health professionals is sustained. Also, PROMs have clear relevance to acute care but their application to domains such as chronic disease and mental illness remains less well-developed.

Chapter 2.3 concerns the ambiguous concept of clinical quality. Most performance measurement schemes consider the outcomes of health care to be a principal focus but their use can be problematic, for example if the outcomes cannot realistically be assessed in a timely or feasible fashion. This is particularly important for chronic diseases. Measures of process then become important signals of future success (Donabedian 1966). Process measures are based on actions or structures known to be associated with health system outcomes in either the health or the responsiveness domains. An example might be appropriate prescribing, known from research evidence to contribute eventually to good outcomes. The concept of effective coverage is an important population health process measure (Shengelia et al. 2003) that seeks to move beyond crude measures of activity in order to adjust for ineffective or inappropriate care. Box 6.1.2 summarizes the basic advantages and disadvantages of outcome and process indicators and the areas of performance measurement for which they are most useful.

Financial protection from catastrophic expenditure associated with ill-health is a fundamental health system concern and has been the driving force behind the systems of universal health insurance enjoyed in most high-income countries. However, the issue remains acute in many lower-income countries that show massive variations in the extent to which households (especially the poor) are protected from catastrophic expenditure. Chapter 2.4 notes that one major challenge is to move beyond the immediate expenditure on health care in order to trace longer-term implications for households' wealth and savings.

Chapter 2.5 shows that work in the responsiveness domain is at an early stage. Patient satisfaction, timely care and respect are important

Box 6.1.2 Usefulness of structural outcome and process indicators

Type of indicator	Advantages	Disadvantages	Areas best used
Outcome indicators	<ul style="list-style-type: none"> ▪ Stakeholders often find outcome measures more meaningful ▪ Direct attention to, and focus on, health goals of the patient ▪ Encourage 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 ▪ Take time to collect ▪ Require large sample sizes to detect statistically significant effects ▪ Can be difficult to measure (i.e. wound infection) 	<ul style="list-style-type: none"> ▪ To measure quality of homogeneous procedures ▪ To measure quality of homogeneous diagnoses with strong links between interventions and outcomes ▪ To measure quality of interventions in heterogeneous populations with a common condition
Process indicators	<ul style="list-style-type: none"> ▪ Easily measured without major bias or error ▪ More sensitive to quality of care ▪ Easier to interpret ▪ Require smaller sample size to detect statistically significant effects ▪ Can often be observed unobtrusively ▪ Provide clear pathways for action ▪ Capture aspects of care that are 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 for patients unless they understand how they relate to outcomes ▪ May be manipulated easily 	<ul style="list-style-type: none"> ▪ To measure quality of care, especially for treatments in which technical skill is relatively unimportant ▪ To measure quality of care of homogeneous conditions in different settings

Source: Adapted from Davies 2005; Mant 2001

issues and many countries are experimenting with patient administered questionnaires but there are very few generally accepted measures of performance that can be compared readily across systems. Measurement of even apparently straightforward concepts such as waiting time has been surprisingly problematic (Sicilliani & Hurst 2005). Development of generally accepted summary measures of responsiveness is therefore a priority for future research.

Equity is a central concern of many health systems and increased use of sample household health surveys in many countries is increasing the potential to develop meaningful performance measures. Chapter 2.6 shows that considerable progress has been made in developing summary measures of equity that permit comparison across health systems and over time. Whenever relying on self-reported health or health-care utilization, a fundamental methodological concern is whether variations are in some sense due to reporting bias. The increased use of electronic health records and objective measures such as biomarkers may help to address this.

Productivity measurement offers an intellectual framework for drawing together the various measures of performance discussed above and relating levels of achievement to the resources consumed. As discussed in Chapter 2.7, this is an immensely challenging undertaking in practice. For example, comparisons between hospitals must allow for patients' different types of case-mix. The challenges become even more daunting in comparisons of health systems. However, there is progress in the methodology for addressing issues of comparability – the notion of adopting disease-based approaches to measure productivity appears especially promising.

Statistical issues

The attribution problem is fundamental when seeking to interpret performance data. This refers to the process of determining what has caused the observed performance and to which practitioners, organizations or agencies any variations in performance should be attributed. It is critical that the causality behind observed measures is attributed to the correct sources in order to inform policy, improve service delivery and ensure accountability. The sensitive nature of correct attribution becomes even more important with increased use of publicly reporting performance data and performance-based payments. Chapter 3.1

stresses that researchers and policy-makers should be careful to control properly for measurement and attribution error when using statistical methods to evaluate causal relationships and inform policy. The key considerations are summarized in Box 6.1.3.

Risk adjustment is the usual approach for addressing the attribution problem. It seeks to adjust performance data to account for variations in patient or population characteristics and can be used for detailed comparison of health-care providers or broad comparisons of population health. Since the early efforts with DRGs in the United States the methods of risk adjustment have been steadily refined over a forty-year period, particularly for adjusting for outcomes for specific diseases or health-care treatments. It is noteworthy that risk adjustment (for co-morbidity, age and other patient risk factors) was central to the New York scheme for public reporting of providers' mortality rates for coronary artery bypass graft surgery.

Chapter 3.2 describes the major progress made within risk adjustment in health care but also highlights many remaining challenges. The key lessons learnt to date are summarized below.

- Optimal risk-adjustment models result from a multidisciplinary effort in which clinicians interact with statisticians as well as experts in information systems and data production.
- Different practice patterns, patient characteristics and data specifications may limit the transferability of models across different countries. Clinicians and methodologists should examine clinical validity and statistical performance before applying a model developed in another setting.
- Decision-makers should be wary of using statistical summary measures (e.g. R-squared values) to draw conclusions about performance on risk-adjustment models as these values may not capture the model's predictive ability for different patient subgroups.
- 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.

A central concern in health care remains the quality (especially completeness) of the data on which risk adjustment is undertaken, especially the presence of co-morbidities or other complications. Recording of these data is ultimately dependent on the practitioners whose per-

Box 6.1.3 Key considerations when addressing causality and attribution bias

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

- Assess existing reports of research investigating a possible causal and attributable link between the agents being assessed and the quality outcome proposed, with particular attention to:
 - the study methodology;
 - its controls for confounding variables; and
 - generalizability of the study sample.
- Undertake prospective analyses to identify critical pathways involved in the achievement of desired and undesired processes and outcomes of care. These analyses should try to identify:
 - possible confounders; and
 - extent to which agents under assessment are/can be clustered into homogeneous groupings.
- In new performance measurement initiatives, carefully consider sources of random and systematic error in measurement and sampling when developing the design. Institutionalize data collection procedures that maximize the reliability and accuracy of data (both primary and secondary) used for quality assessment.
- Employ risk adjustment techniques when evaluating the relationship between agents under assessment and the quality indicators. Use hierarchical models to account for the clustering of data within different levels of the health system under analysis. Consider using statistical methods such as propensity scores or instrumental variables.
- Causality and attribution bias cannot be eliminated completely, even when utilizing best practice. Monitor carefully any unintended impacts from biases in assessment of performance, especially when reimbursement or other incentives are linked to the measures.

Source: Adapted from Terris & Aron 2009

formance is being assessed, with an ever-present threat to the integrity of the data if the incentives associated with performance comparison are too stark. Furthermore, most risk adjustment efforts are still work in progress. Consequently, there is often a need for careful qualitative clinical commentary on any risk-adjusted data as technical limitations are common. However, as risk adjustment is almost always essential for performance measurement to secure credibility with practitioners it is important to sustain efforts to improve current methodologies.

In the public health domain, it is key that risk adjustment establishes what the agency under scrutiny is accountable for. For example, in the short run a health system has to deal with inherited epidemiological patterns and risky behaviours. This implies a major need for risk adjustment when comparing different health systems. In the longer run, the health system might be expected to be accountable for improving epidemiological patterns and health-related behaviour. This changes the nature of risk adjustment as the health system can now be considered accountable for at least some of the underlying causes of measured outcomes.

Proper statistical treatment of performance indicators is essential if appropriate policy inferences are to be drawn, given the large degree of random variation present in most performance indicators. The Royal Statistical Society has produced a protocol that seeks to summarize best practice for the selection, collection, analysis and presentation of performance data (Bird et al. 2005). One of the most important issues is the need to present measures of uncertainty (e.g. confidence intervals) alongside any performance measure. For example, it is noteworthy that confidence intervals were a central feature of the New York cardiac surgery public reporting initiative. The intention is to signal when a variation in performance is a matter for concern and the potential urgency of the need for intervention. Chapter 3.3 gives a particular example of how this approach can be applied – statistical control charts track a provider's performance over time and identify in a timely fashion any systematic deviation from expected levels of attainment. The authors discuss the criteria for selecting statistical methods including utility, verity, simplicity and responsiveness. These can be applied to most statistical methods of analysis applied to performance measurement although any choice involves some trade-offs. However, improved treatment of uncertainty is essential if performance measures are to retain credibility with patients, professionals and regulators.

Chapter 1.1 highlights the many different uses and users of performance measurement in the health system – it will often be the case that different levels of aggregation of performance measures will be needed for different uses. Chapter 3.4 discusses the role of composite measures of performance of whole systems and organizations. The chapter highlights the considerable controversy that exists as the science of composite measurement is still embryonic. However, while composite measures are of questionable direct use to patients or professionals they serve as a crucial element in promoting accountability to legislatures, governments and citizens in general. It is therefore important that (to the extent that data permit) they are credible; are constructed using transparent methods; and that users of composite indicators, including the media, are made aware of their limitations.

The choice of weights (or importance) attached to the component measures is fundamental to composite indicators. All the evidence suggests that individual citizens attach widely varying importance to aspects of performance. This indicates that the choice of weights is first and foremost a political undertaking, requiring the decision-maker to have political legitimacy. Analysis can therefore inform but should not determine the choice of weights. The body of economic methodology for inferring weights includes methods for calculating willingness to pay valuations or to elicit patients' preferences from rankings of alternative scenarios or direct choice experiments. However, these have not been applied widely to the construction of composite indicators of health system performance (Smith 2002). Box 6.1.4 summarizes the advantages and disadvantages of using composite indicators for health performance assessment.

Incentives and performance information

Accountability is not just about the production of performance measures, it also requires mechanisms with which to hold agents to account. In other words, the agent needs an incentive to take notice of the performance measures. For example, it is noteworthy that important comparative data on hospital performance in Scotland (including risk adjusted mortality rates in various specialties) were routinely fed back to hospital boards, albeit without any deliberate publicity. However, they appear to have had little impact on boards or clinicians and many senior managers and physicians claimed to have no knowl-

Box 6.1.4 Advantages and disadvantages of composite indicators

Advantages

- Offer a broad assessment of system performance.
- Place system performance at centre of the policy arena.
- Enable judgment and cross-country comparison of health system efficiency.
- Offer policy-makers at all levels the opportunity to set priorities and seek out performance improvement in these areas.
- Clearly indicate which systems represent the best overall performance and improvement efforts.
- Can stimulate better data collection and analytical efforts across health systems and nations.

Disadvantages

- Composite indicators may disguise failings in specific parts of the health-care system.
- Composite indicators make it difficult to determine where poor performance is occurring and consequently may make policy and planning more difficult and less effective.
- Indicators often have high positive correlation which can lead to double counting.
- In seeking to cover many areas, composite indicators may use feeble data that may also question the methodological soundness of the entire indicator.
- Aggregation of the data may conceal contentious individual measures within the composites.
- Composite indicators may ignore dimensions of performance that are difficult to measure, leading to adverse behavioural effects.
- Methodology on applying weights to composite indicators is not adequately developed and may reflect only certain preferences.

Source: Adapted from Smith 2002

edge of the data (Mannion & Goddard 2001). In contrast, the star ratings report cards prepared for English hospitals had a profound impact on behaviour because the data were publicly reported and had some very real incentives and sanctions attached to them, for the organizations and for the senior managers.

Incentives can arise as an incidental by-product of other system reforms. For example, almost all finance mechanisms introduce acci-

dental incentives that may be benign or indeed reinforce a desire to secure improved performance. Accidental incentives can also lead to adverse consequences. Performance data can often fulfil an important role in correcting adverse incentives – for example, careful monitoring of performance can abate the common incentive to skimp on quality of care that results from hospital case payment (DRGs). However, performance measurement itself can give rise to unintended outcomes especially when explicit incentives are attached. Part 5 of the book explores the role of performance incentives under a number of headings: performance targets; public performance reporting; direct financial incentives; and professional improvement.

Health system targets are a specific type of performance measurement and incentive scheme. They comprise a quantitative expression of an objective to be met in the future. Brought to health policy from the business world, the main idea is that more organized and efficient efforts will be made to meet goals that are defined explicitly as targets. Targets are expected to be SMART – specific, measurable, accurate, realistic and time bound (van Herten & Gunning-Schepers 2000). The governments of many countries (including United States, United Kingdom, European Member States, Australia, New Zealand) have experimented with targets in health care.

Health system targets have traditionally been used extensively in public health but Chapter 5.1 indicates that reports of measurable success are rare. The English experience with the 1992 *Health of the Nation* strategy is typical. Based on WHO's *Health for All* initiative this 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 policymaking was negligible' (Department of Health 1998). Hunter (2002) summarizes its failings under six broad headings:

1. Appeared to be a lack of leadership in the national government.
2. Policy failed to address the underlying social and structural determinants of health.
3. Targets were not always credible and were not formulated at a local level.
4. Poor communication of the strategy beyond the health system.
5. Strategy was not sustained.
6. Partnership between agencies was not encouraged.

In the past decade, targets have featured especially strongly in English health-care policy. Starting in 1998 the Treasury issued strategic targets (PSAs) to all government departments including the health ministry (Smith 2007). PSAs 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 star rating report cards 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 (Bevan & Hood 2006). This success can be attributed to the following characteristics.

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

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

Targets provide a straightforward way of highlighting key objectives and can be very successful if designed and implemented correctly. However, some of the notable risks associated with their use are summarized in Box 6.1.5. The conclusions from this experience indicate that performance targets offer some scope for focusing system attention on specific areas of endeavour but are unlikely to secure performance improvement unless they are implemented carefully alongside other improvement initiatives such as more general inspection and regulation.

Public performance reporting is established within health care and is congruent with an increasing broader trend for transparency in society. Even if it had no discernable impact on health system performance, it would be necessitated by growing public demand for important outcome information to be made available to patients and the pub-

Box 6.1.5 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.
- Excessively aggressive targets may undermine the reliability of the data on which they are based.
- Excessively aggressive targets may induce gaming or other undesirable behavioural responses.
- Targets may encourage a narrow, mercenary attitude rather than altruistic professional motivation.
- Targets require continual monitoring and updating to verify that they remain relevant and are not undermined by professional interests.

Source: Smith 2008

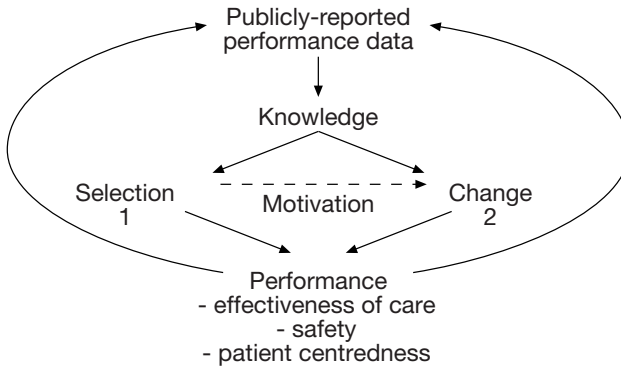


Fig. 6.1.1 Pathways for improving performance through publicly reported data

Source: Berwick et al. 2003

lic, both to enhance public accountability and to inform health-care consumers. Moreover, public reporting can improve quality through two pathways (Fig. 6.1.1): (i) selection pathway by which consumers become better informed and select providers of higher quality; and

(ii) change pathway by which information helps providers to identify areas of underperformance and thus acts as a stimulus to improve.

Chapter 5.2 examines the growing experience of placing information in the public domain and reports considerable evidence that publication of provider performance measures leads to performance improvement (Hibbard et al. 2005). Although the immediate purpose has often been to facilitate and inform patient choice, there is little evidence that patients make direct use of report cards. However, report cards do appear to promote performance improvements in providers by means of their impact on reputation – the change pathway. For example, long-standing use of coronary artery bypass graft report cards in two American states (New York and Pennsylvania) has unequivocally been associated with improvements in risk-adjusted mortality. Nevertheless, there is continuing debate about whether these results necessarily imply that they have been beneficial and a number of adverse outcomes associated with the schemes have also been reported (Dranove et al. 2003; Schneider & Epstein 1996).

- Coronary artery bypass graft 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).
- Initiative has increased Medicare expenditures with only a small improvement in population health.
- Practitioners were concerned about the absence of quality indicators other than mortality; inadequate risk adjustment; and unreliability of data provided by physicians and hospitals.

This experience underlines the importance of carefully monitoring and evaluating the outcomes of incentive schemes, such as public reporting. It suggests several points that should be taken into account when implementing public disclosure of data.

- Give careful consideration to the purpose of the disclosure and the type of information that different stakeholders of the health system will want and are able to use.
- Give careful consideration to the impact on quality of care that may result from public disclosure of information. Where appropriate, public disclosure of information should be integrated with other quality improvement strategies (Marshall et al. 2000).

- Enhance the credibility and usefulness of public performance reports by creating them in collaboration with physicians and other legitimate interest groups (Marshall et al. 2000; Schneider & Epstein 1996).
- When reporting data, implement careful risk adjustment in order to assure legitimacy of the scheme with providers and offer accurate comparisons between outcomes and providers (Iezzoni 2009; Marshall et al. 2000). Detailed information on the risk adjustment strategies used should be made available for public scrutiny alongside the reported information.

There is no doubt that clinicians and other actors in the health system generally do respond to financial incentives (Dudley 2005). When performance measurement is incorporated into financial incentive regimes it offers a potentially 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 offered through systems of accreditation that offer rewards in the form of access to markets or extra payments for meeting structural requirements. Germany has an accreditation system of this sort at the regional level in which specific quality indicators are used for accreditation (www.G-BA.de).

Yet accreditation is a very blunt incentive instrument and Chapter 5.4 summarizes the evidence that policy is shifting towards very much more direct and focused incentives, for individuals and for organizations. The author indicates that most experiments so far have been small scale and it is difficult to draw strong conclusions as the results are difficult to assess with any confidence. Notwithstanding, there is evident need for increased experimentation and research.

Chapter 5.4 also highlights the many issues that need to be considered when designing performance incentive schemes, summarized in Box 6.1.6 below. This complexity is one reason why it is difficult to make a definitive evaluation of financial incentives. Furthermore, translation of results from one institutional setting to another must be treated with caution because other health system instruments (e.g. financing mechanism) may interact with the incentives to produce unexpected results. In short, any scheme requires constant monitoring to ensure that there are no unintended responses to incentives (e.g. cream-skimming, gaming); that the incentive scheme does not jeopardize the reliability of the performance data on which it relies; and that it does not compromise unrewarded aspects of performance.

Box 6.1.6 Design issues for pay-for-performance schemes

The behaviour induced by a financial incentive varies according to its design along each of the following dimensions (discussed in more detail in Chapter 5.4).

Organizational vs. individual

Financial incentives can be awarded at either group or individual level. A group-level payment may be less likely to motivate individual behaviour and may encourage free-riding but may be better at inducing cooperation and coordination. The incentive should be offered at the level at which it most directly motivates the party responsible for the action being incentivized.

Absolute vs. relative

Incentives based on relative performance measures may increase competition amongst providers, especially if this information is made publicly available. Absolute incentive payments offer the payee more certainty about the attainment of payment and may increase motivation.

Short term vs. long term

Long-term payments can lead to greater investment in structural change and processes of care and provide a longer time frame in which to observe results. Short-term payments may have the benefit of appearing more salient and corresponding more closely to the action being incentivized. They may also impose greater administrative costs and encourage more myopic behaviour.

Reward vs. penalty

Incentives structured as rewards or penalties (e.g. withholding payment) may affect providers' attitudes towards performance and therefore have differential incentive effects.

Size and power of payment

It is important to ensure that an incentive payment is large enough to cover the marginal costs involved in adjusting behaviour to achieve the targeted results. However, payment levels should not be higher than the level required to encourage participation in the incentive scheme. Furthermore, the link between performance and reward (power of the incentive) needs to be calibrated carefully.

Box 6.1.6 cont'd**Choice of performance measures**

The best performance measures on which to attach financial incentives are those which lie within the control of the physician. Usually these are structural or process of care indicators. Outcome indicators are more likely to be influenced by external factors and are therefore less favourable.

Source: Conrad 2009

Another important use of performance measurement is to provide clinical practitioners with feedback on their performance relative to their peers, with the intention of stimulating performance improvement. Databases serving this purpose exist in many countries. For example, providers in Sweden contribute to quality registers by voluntarily collecting individual-based data on patient characteristics, diagnosis, treatments, experiences and outcome that are shared with other members of the register. The quality registers have an explicit aim to facilitate the improvement of quality in clinical work through continuous learning and development (Rehnqvist 2002). Indeed there is a strong argument that performance measurement should become an inherent element in a professional's lifelong learning. This suggests the need for a prominent role for performance measurement principles in early clinical training.

There is much debate on whether information for professional improvement should be anonymized or made available to the public. Evidence suggests that such performance measurement schemes need to be designed and owned by the professionals who use them in order to be effective (Rowan & Black 2000). It is argued that the most constructive systems are those that encourage positive and cooperative behaviour amongst practitioners and avoid public threats to their professional or commercial standing. The latter may encourage defensive behaviour that could lead to gaming or cream-skimming. Indicators used for professional improvement should therefore:

- reflect meaningful aspects of clinical practice with strong scientific underpinning;
- assure close risk adjustment of indicators;

- allow exclusions of certain patients, e.g. those who refuse to comply with treatment;
- facilitate interpretability;
- represent services under a provider's control;
- assure high accuracy;
- minimize cost and burden.

Furthermore, it is important to measure not only the outcomes of care but also 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. Experience from Sweden and elsewhere (e.g. Netherlands, Denmark) suggests that public and professional needs can be reconciled. For example, quality registers such as the Danish National Indicator Project (www.nip.dk) publish outcomes on individual practitioners. In any case, it is likely that patient advocacy groups will increasingly demand that more performance data should be made available. The challenge for the professions is to ensure that this trend is harnessed for good rather than leading to defensive professional behaviour. One solution lies in careful development of acceptable statistical risk adjustment schemes and careful presentation of statistical data so that the public and media are better equipped to understand and interpret the information made available to them.

Politics of performance measurement

It is inevitable that performance information of any power creates winners and losers. A recurring theme throughout the book is therefore the immensely political nature of any attempt to measure performance within the health system. Inspired by pioneers such as Florence Nightingale and Ernest Codman, the very earliest efforts to measure performance were ultimately frustrated by the opposition of elements within the medical profession and a lack of resolve amongst politicians (Spiegelhalter 1999). One hundred and fifty years later, the earliest opposition to Nightingale's proposals to measure surgical outcomes still sounds remarkably familiar to contemporary readers.

The political nature of performance measurement is an inevitable consequence of its power to challenge vested interests within the health system. There is an enormous range of interest groups, often encompassing (amongst others):

- taxpayers
- voters
- patient groups
- clinical professionals
- insurers and other purchaser organizations
- provider organizations
- pharmaceutical companies
- governments
- geographical interests
- age groups
- social groups (income, ethnicity).

Performance information often serves the interests of some of these groups but will also challenge others. The natural response of those under challenge will be to contest the veracity, completeness and relevance of the information provided. High-quality statistical analysis of data, such as risk adjustment, is therefore imperative to assure the credibility of the performance measurement.

Chapter 1.1 argues that performance information plays a prime role in enabling principals to hold agents to account more effectively within the health system. Transparency, in the form of performance information, is a fundamental requirement for enhancing the accountability of governments, provider organizations, professionals and insurers to patients and the broader citizenry. Furthermore, as summarized in Chapter 5.2 on public reporting, many authors have argued that enhanced accountability leads to improved performance in a virtuous improvement circle.

If performance indicators are to promote accountability they must address the specific questions of each discrete audience and be presented with appropriate clarity that resonates with the various constituencies. These may include patients, the broader public, professionals, the media and researchers. A key requirement in the accountability cycle is to identify the targets and shape the analysis and presentation of the measures to suit their needs.

In most health systems the many actors formally charged with governance of institutions and professions form an especially important constituency. These might include the boards of governors of provider organizations; professional conduct committees; a wide range of regulators; and elected representatives in local and central government. Performance information plays a particularly important role in enabling these constituencies to discharge their roles effectively. Comparative performance information should be an important resource and all those charged with governance should be given the capacity to demand and understand such data.

One specific issue highlighted by Busse and Smith is performance measurement's potential to undermine the traditional approach of clinical professionalism that encourages clinicians to do the best for their patients, regardless of pecuniary or other incentives. Pursued to excess, reliance on a limited range of specific indicators may distort professional behaviour by encouraging treating to the test – concentrating on measured aspects of care at the expense of the unmeasured.

This argument has some force when applied to a system with very partial or distorted performance information but in our view does not compromise the argument for performance measurement. Rather it suggests the need for redoubled efforts to broaden the scope of measurement; to shift from measuring processes to measuring health outcomes wherever possible; to improve the quality of statistical analysis; to ensure that incentives are not distortionary; and to ensure that performance data are used constructively to help professionals to improve the care they offer.

The political element of performance measurement will always exist and therefore one of the fundamental roles of governments will be to nurture informed political debate. This includes ensuring that legitimate interests are empowered to make their case using the best available performance information and that the information is fit for purpose. In particular, it is important to ensure that key constituencies such as the public, patients and professionals are fully engaged in the development, analysis and interpretation of performance measures.

Stewardship perspective on performance measurement

Governments play a major stewardship role in harnessing the full

potential of performance measurement to improve the health system. *The world health report 2000* defined stewardship as: ‘...defining the vision and direction of health policy, exerting influence through regulation and advocacy, and collecting and using information’ (WHO 2000). This summary seeks to outline how performance measurement can help governments to fulfil each of these tasks. We argue that performance measurement offers major opportunities to secure performance improvement and that no health system can be steered adequately 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 a government directly in the formulation and evaluation of policy and in undertaking regulation. However, government’s broader stewardship role is to assure that the necessary flows of information are available, functioning properly and aligned with the design of the health system. Performance measurement is a public good that will not occur naturally and therefore government has the fundamental role of ensuring that maximum benefit is secured from performance measurement, whether through law, regulation, coordination or persuasion. Implementation requires sustained political and professional leadership at the highest level and assurance that the necessary analytical capacity is available throughout the health system.

Some of the stewardship responsibilities of government are summarized in Box 6.1.7. While these functions and tasks must be in place, government itself is not necessarily required to perform them.

Box 6.1.7 Stewardship responsibilities associated with performance measurement

Development of a clear conceptual framework and a clear vision of the purpose of the performance measurement system:

- align with accountability relationships
- align with other health system mechanisms (e.g. finance, market structure, IT)

Design of data collection mechanisms:

- detailed specification of individual indicators
- alignment with international best practice

Box 6.1.7 cont'd

Information governance:

- data audit and quality control
- assuring public trust in information
- assuring well-informed public debate

Development of analytical devices and capacity to help understand the data:

- commissioning appropriate research on (e.g.) risk adjustment, uncertainty, data feedback mechanisms
- ensuring analysis is undertaken efficiently and effectively
- ensuring local decision-makers understand the analysis

Development of appropriate data aggregation and presentational methods:

- ensuring information has appropriate impact on all parties
- mandating public release of summary comparative information
- ensuring comparability and consistency

Design of incentives to act on performance measures:

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

Proper evaluation of performance measurement instruments:

- ensuring money is spent cost effectively on information resources

Managing the political process

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

Future priorities

Given increasing demand and the wide 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 suggests that any policy development should take account of the following recommendations.

1. Develop a clear conceptual framework and a clear vision of the purpose of the performance measurement system in alignment with the accountability relationships inherent in the health system.
2. Ensure that definitions of performance indicators are clear and consistent and fit the chosen conceptual framework.
3. Indicators should:
 - aim to measure what matters, specifically to: promote health, improve patient care and ensure prudent utilization of health system resources;
 - be statistically sound and presented in ways that are straightforward to interpret in order to reduce the likelihood of manipulation or misinterpretation;
 - fully acknowledge any data limitations, including levels of uncertainty and lack of timeliness.
4. Pay more attention to improving the comprehensibility and utility of performance data, particularly how to improve its interpretation by patients, providers and practitioners.
5. Enhance managers' and clinicians' capacity to understand and use information. Use of performance data should become an intrinsic part of clinical education and lifelong professional development.
6. Incentives to act upon performance measures should be designed carefully. Monitor closely how performance information impacts on behaviour and take action to enhance beneficial outcomes and negate adverse consequences.
7. Policy-makers should pay particular attention to the broader health system, ensuring that performance measurement is aligned with the design of mechanisms such as finance and market structures and recognizing the organizational context within which performance data are collected and disseminated.
8. Performance measurement systems should be monitored frequently and evaluated to identify opportunities for updating and improvement and any unintended side effects.

9. Ensure effective management of the political process of performance measurement. Amongst other things, encourage healthy political debate and ensure that specific interest groups do not capture the performance information system.

While arguing very strongly for increased use of performance measurement throughout the health system we recognize that this is a costly undertaking that diverts valuable resources from health services. It is imperative that all performance measurement initiatives are undertaken effectively and justified with the same cost-effective criteria that should be applied to more conventional health technologies. Many performance measurement initiatives will involve relatively low-cost capture of data that are already required to assure the delivery of high-quality services. However, their utility should be evaluated rigorously when they do involve significant additional costs.

The effectiveness of any performance measurement instrument ultimately should be evaluated not in relation to statistical properties (e.g. accuracy and validity) but more broadly – by the extent to which it promotes or compromises health system objectives. Effective performance measurement alone is not enough to ensure performance improvement – the functions of analysis and interpretation of performance data are also crucial. Furthermore, performance measurement is only one (albeit very important) instrument for securing system improvement. For maximum effect it needs to be aligned with other levers for system reform such as financing, market structure, accountability arrangements and regulation. Without careful attention to these broader health system considerations the performance measurement system will be ineffective.

Health systems are in the early days of performance measurement and there is still huge potential to improve the effectiveness of measurement systems. However, performance measurement offers scope for major health system improvements. Advances in technology are likely to increase this potential still further and 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 within the health system; implement initiatives of proven effectiveness; undertake careful trials of less established mechanisms; and monitor and update performance measurement systems as new knowledge and capacity emerge.

References

- Berwick, DM. James, B. Coye, MJ (2003). 'Connections between quality measurement and improvement.' *Medical Care*, 41(1): I30–I38.
- Bevan, G. Hood, C (2006). 'Have targets improved performance in the English NHS?' *British Medical Journal*, 332(7538): 419.
- Bird, S. Cox, D. Farewell, VT. Golstein, H. Holt, T. Smith, PC (2005). 'Performance indicators: good, bad, and ugly.' *Journal of the Royal Statistical Society*, 168(1): 1–27.
- Conrad, D (2009). Incentives for health-care performance improvement. In: Smith, PC. Mossialos, E. Papanicolas, I. Leatherman, S (eds.). *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge: Cambridge University Press.
- Davies, H (2005). *Measuring and reporting the quality of health care: issues and evidence from the international research literature*. NHS Quality Improvement Scotland.
- Department of Health (1998). *Health of the Nation: a policy assessed*. London.
- Donabedian, A (1966). 'Evaluating the quality of medical care.' *Milbank Memorial Fund Quarterly*, 44(3): 166–206.
- Dranove, D. Kessler, M. McClellan, M. Satterthwaite, M (2003). 'Is more information better? The effects of 'report cards' on health care providers.' *Journal of Political Economy*, 111(3): 555–588.
- Dudley, RA (2005). 'Pay-for-performance research: how to learn what clinicians and policy makers need to know.' *Journal of the American Medical Association*, 283(14): 134–148.
- Hibbard, JH. Stockard, J. Tusler, M (2005). 'Hospital performance reports: impact on quality, market share, and reputation.' *Health Affairs*, 24(4): 1150–1160.
- Holland, WW (1988). *European Community atlas of 'avoidable death'*. Oxford: Oxford Medical Publications (Commission of the European Communities Health Services Research Series No. 3).
- Hunter, DJ (2002). England. In: Marinker, M. McKee, M (eds.). *Health targets in Europe: policy, progress and promise*. London: BMJ Books.
- Iezzoni, L (2009). Risk adjustment for performance measurement. In: Smith, PC. Mossialos, E. Papanicolas, I. Leatherman, S (eds.). *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge: Cambridge University Press.
- Mannion, R. Goddard, M (2001). 'Impact of published clinical outcomes data: case study in NHS hospital trusts.' *British Medical Journal*, 322(7307): 260–263.

- Mant, J (2001). 'Process versus outcome indicators in the assessment of quality of health care.' *International Journal for Quality in Health Care*, 13(6): 475–480.
- Marshall, MN, Shekelle, PG, Leatherman, S, Brook, RH (2000). 'Public disclosure of performance data: learning from the US experience.' *Quality in Health Care*, 8(1): 53–57.
- Nolte, E, McKee, M (2004). *Does health care save lives? Avoidable mortality revisited*. London: The Nuffield Trust.
- Power, M (1999). *The audit society: rituals of verification*. Oxford: Oxford University Press.
- Rehnqvist, N (2002). 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.
- Rowan, K, Black, N. (2000). 'A bottom-up approach to performance indicators through clinician networks.' *Health Care UK*, Spring: 42–46.
- Schneider, EC, Epstein, AM (1996). 'Influence of cardiac-surgery performance reports on referral practices and access to care.' *New England Journal of Medicine*, 335(4): 251–256.
- Sequist, T, Bates, D (2009). Developing information technology capacity for performance measurement. In: Smith, PC, Mossialos, E, Papanicolas, I, Leatherman, S (eds.). *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge: Cambridge University Press.
- Shengelia, B, Murray, CJL, Adams, OB (2003). Beyond access and utilization: defining and measuring health system coverage. In: Murray, CJL, Evans, DB. *Health systems performance assessment: debates, methods and empiricism*. Geneva: World Health Organization.
- Sicilliani, L, Hurst, J (2005). 'Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries.' *Health Policy*, 72(2): 201–215.
- Smith, PC (2002). 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.
- Smith, PC (2007). Performance budgeting in England: experience with Public Service Agreements. In: Robinson, M (ed.). *Performance budgeting: linking funding and results*. Washington DC: International Monetary Fund.
- Smith, PC (2008). England: intended and unintended effects. In: Wismar, M, McKee, M, Ernst, K, Srivastava, D, Busse, R (eds.). *Health targets*

- in Europe: learning from experience*. Copenhagen: WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies.
- Spiegelhalter, DJ (1999). 'Surgical audit: statistical lessons from Nightingale and Codman.' *Journal of the Royal Statistical Society*, 162(1): 45–58.
- Terris, DD. Aron, DC (2009). Attribution and causality in health-care performance measurement. In: Smith, PC. Mossialos, E. Papanicolas, I. Leatherman, S (eds.). *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge: Cambridge University Press.
- van Herten, L. Gunning-Schepers, LJ (2000). 'Targets as a tool in health policy. Part I: lessons learned.' *Health Policy*, 53(1): 1–11.
- WHO (2000). *The world health report 2000*. Geneva: World Health Organization.