



# Euro Observer

The Health Policy Bulletin  
of the European Observatory on Health Systems and Policies

Spring 2008

Volume 10, Number 1

## Contents

Principles of performance measurement	1
Long term care quality monitoring using the interRAI Common Clinical Assessment Language	5
Public reporting of performance in the USA: The New York State CSRS	7
Using composite indicators to measure performance in health care	8
Attribution and causality bias in health care performance measurement	10

## Principles of performance measurement

Irene Papanicolas, Peter C Smith and Elias Mossialos

Information plays a central role in the ability of a health system to deliver effective health care and secure population health. It is used for a variety of different purposes in health care: to secure accountability within the system, to determine appropriate treatments for patients, to facilitate patient choice and/or for managerial control. Information also plays an important broader role in assuring governance of the health system. Early efforts at health system performance measurement can be dated as far back as 250 years ago.<sup>1</sup> Later, in the 1800s and early twentieth century, Florence Nightingale and Ernest Codman advocated using systematically collected data to inform and improve performance. Yet, because of insuperable professional, practical, and political barriers performance policies were never generally applied.

It is only in the past few decades that sustained developments in health system performance measurement and assessment have emerged. On the demand side, citizens require increasing accountability from health care professions and institutions and want to make more informed choices as patients.<sup>2</sup> On the supply side, advances in information technology have made it much cheaper and easier to collect and process data.

However, despite the increasing use of performance measurement tools in developed health systems, questions still remain about the best way to collect, disseminate and use performance data efficiently. This issue of

*Euro Observer* will explore some of the issues more closely through case studies on composite indicators in performance measurement, attribution and causality, and the use of performance measurement in long term care and cardiac surgery. Here, we provide a brief overview of the principles of performance measurement, highlighting some of the main policy implications.

### Defining and measuring performance

Performance measurement evaluates the extent to which a health system meets its key objectives. These objectives reflect different historical trajectories, political, financial and organizational priorities and the power of interest groups and stakeholders. The World Health Report 2000 defined three intrinsic goals of health systems - improving health, increasing responsiveness to the legitimate demands of the population, and ensuring that financial burdens are distributed fairly.<sup>3</sup>

Generally, 'health' is taken to represent both health outcomes from health care and improvements in general health status. 'Responsiveness' captures dimensions 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. Improvements in responsiveness depend on a health system's ability to deliver these dimensions more effectively. 'Productivity' refers to notions of efficiency



WHO European Ministerial  
Conference on Health Systems:  
"HEALTH SYSTEMS,  
HEALTH AND WEALTH"  
Tallinn, Estonia, 25-27 June 2008

This issue of *Euro Observer* has been written for the WHO European Ministerial Conference on Health Systems, 25-27 June 2008, Tallinn, Estonia. The articles are based on a study, *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects*, which has been commissioned for the Conference and will be published by WHO and Cambridge University Press. For more information on the Ministerial Conference go to: <http://www.euro.who.int/healthsystems2008>

The Observatory is a partnership between the WHO Regional Office for Europe, the Governments of Belgium, Finland, Greece, Norway, Spain, Slovenia and Sweden, the Veneto Region of Italy, the European Investment Bank, Open Society Institute, the London School of Economics and Political Science and the London School of Hygiene & Tropical Medicine.

**Table 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.	<ul style="list-style-type: none"> <li>• Life expectancy</li> <li>• Healthy life expectancy</li> <li>• Years of life lost</li> <li>• Avoidable mortality</li> <li>• Disability Adjusted Life Years</li> </ul>
Individual health outcomes	Measures of individuals' 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"> <li>• Short form 36 (SF-36)</li> <li>• EQ5D</li> </ul> Disease-specific measures: <ul style="list-style-type: none"> <li>• Arthritis Impact Measurement Scale</li> <li>• Parkinson's Disease Questionnaire (PD-39)</li> </ul>
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 whether these actions are carried out in a technologically sound manner.	Outcome measures: <ul style="list-style-type: none"> <li>• Health status</li> <li>• Specific post-operative mortality rates</li> </ul> Process measures: <ul style="list-style-type: none"> <li>• Frequency of blood pressure measurement</li> </ul>
Responsiveness of health system	Measures of the way individuals are treated and the environment in which they are treated during health system interactions.  Concerned with issues of patient dignity, autonomy, confidentiality, communication, prompt attention, social support and quality of basic amenities.	<ul style="list-style-type: none"> <li>• Patient experience measures</li> <li>• Patient satisfaction measures</li> <li>• Indicators from broader population surveys</li> </ul>
Equity	Measures of the extent to which there is equity in health, access to health care, responsiveness and financing.	<ul style="list-style-type: none"> <li>• Utilization measures</li> <li>• Rates of access</li> <li>• Use-needs ratios</li> <li>• Spending thresholds</li> </ul>
Productivity	Measures of the productivity of the health care system, health care organizations and individual practitioners.	<ul style="list-style-type: none"> <li>• Labour productivity</li> <li>• Cost effectiveness measures (i.e. for treatments and interventions)</li> <li>• Technical efficiency (measures of output/input)</li> <li>• Allocative efficiency (i.e. comparing different ways of improving health – cost utility analysis)</li> </ul>

and effectiveness. In each of these three domains both the average attainment of a health care system and its distribution across the population (equity) are of interest.

Table 1 summarizes the main aspects of performance measurement that are used to evaluate health systems.

### Information needs and requirements

Modern health systems are complex entities with many different stakeholders such as patients, health care professionals, health care provider and purchaser organizations, regulators and central and local governments. Each of these groups has different needs in terms of the nature of information and the detail, timeliness and level of aggregation they require. Therefore, a fundamental challenge in performance measurement is how to design systems that cater for diverse needs (see Table 2).

### Uses of performance data

There have been various efforts over the last two decades to use large-scale information for performance assessment and improvement. Early initiatives concentrated on collecting and organizing such data for management purposes to assist with cost-containment efforts and resource allocation, as was the case in England during the 1980s. Later, large databases were also used to assist with evidence-based decision making in health planning and accountability, as in Canada.

Increasingly, data has been placed in the public domain to provide citizens with information on provider performance. 'Report cards' take the form of aggregated summaries of performance indicators such as waiting times, satisfaction ratings, and risk adjusted post-operative mortality rates. By enhancing accountability and stimulating provider improvement, these initiatives can secure important gains. However, to date little use of such data has been made by citizens, either as patients or payers.<sup>4</sup> Moreover, public reporting may lead to other adverse outcomes. For example, the focus on specific performance measures, such as waiting times, may draw attention away from other areas that are not reported. And there is also some evidence that if not designed carefully public reporting can lead to cream-skimming or gaming by providers.<sup>5</sup>

Performance measurement also has been used in conjunction with explicit financial incentives to

**Table 2: Information Requirements for Health System Stakeholders**

Stakeholder	Information needs	Data Requirements
Government	<p>Depending on the role of the government in health care, needs may include:</p> <p>Monitoring of:</p> <ul style="list-style-type: none"> <li>regulatory procedures;</li> <li>financing procedures;</li> <li>information collection;</li> <li>regulatory effectiveness and efficiency.</li> </ul>	<p>Information on:</p> <ul style="list-style-type: none"> <li>performance at national and international levels;</li> <li>access and equity of care;</li> <li>utilization of service and waiting times.</li> </ul>
Regulators	<ul style="list-style-type: none"> <li>To protect patients' safety and welfare;</li> <li>To ensure the market is functioning efficiently.</li> </ul>	<p>Timely, reliable and continuous information on patient safety and welfare.</p>
Payers (taxpayers or members of insurance funds)	<ul style="list-style-type: none"> <li>To ensure money is being spent effectively, efficiently and in line with payer's expectations.</li> </ul>	<p>Aggregate, comparative performance measures;</p> <p>Information on:</p> <ul style="list-style-type: none"> <li>productivity and cost-effectiveness;</li> <li>access to and equity of care.</li> </ul>
Purchaser organizations	<ul style="list-style-type: none"> <li>To ensure that the contracts they offer their patients are in line with the objectives their patients and payers expect.</li> </ul>	<p>Information on</p> <ul style="list-style-type: none"> <li>patient experiences and patient satisfaction;</li> <li>provider performance;</li> <li>cost-effectiveness of treatments.</li> </ul>
Provider organizations	<ul style="list-style-type: none"> <li>To monitor and improve existing services;</li> <li>To assess local needs.</li> </ul>	<p>Aggregate clinical performance data;</p> <p>Information on:</p> <ul style="list-style-type: none"> <li>patient experiences and patient satisfaction;</li> <li>access and equity of care;</li> <li>utilization of service and waiting times.</li> </ul>
Physicians	<ul style="list-style-type: none"> <li>To stay up to date with current practice;</li> <li>To be able to improve performance.</li> </ul>	<p>Information on:</p> <ul style="list-style-type: none"> <li>current practice, and best practice;</li> <li>comparative performance.</li> </ul>
Patients	<ul style="list-style-type: none"> <li>Ability to make a choice of provider when in need;</li> <li>Assurance of good emergency care.</li> </ul>	<p>Information on:</p> <ul style="list-style-type: none"> <li>location and quality of nearby emergency health services;</li> <li>quality of options for elective care.</li> </ul>

reward provider performance.<sup>6</sup> While research shows that clinicians do respond to financial incentives, little compelling evidence exists on the effects of financial incentives rewarded for improvements in quality of care. Such incentives take the form of payment-for-performance schemes or direct quality improvement payments. However, as experiments to date have been small, and often coupled with other incentive payments, conclusions have been mixed and difficult to draw with confidence. Many issues need to be resolved both in design and implementation of these schemes, as well as their evaluation. Yet, the level of detail that performance measurement can report on suggests that such policies are feasible, and could very well be promising.

Health 'targets' constitute a specific type of performance measurement and incentive scheme. These are a quantitative expression of an objective to be met in the future. If well designed, targets can be very useful in stimulating progress towards the achievement of specific goals. However, targets are selective and focus on specific areas, running the risk of leaving untargeted areas neglected.<sup>7</sup> As Goodhart<sup>8</sup> emphasised 'any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes' and therefore existing targets should be routinely scrutinised for their continued relevance and effectiveness.

Performance measurement can also be used to provide feedback to clinical practitioners on their actions and how these compare to those of their peers. For example, databases exist in Sweden in the form of 'quality registers' where clinical information is available for the use of professionals. In this context, there is much debate on whether information should be anonymized and/or made available to the public, and some concerns over potential gaming or other unintended behavioural effects (such as cream skimming) on the part of providers. Overall, these databases require careful statistical risk adjustment to control for confounding patient characteristics in order for the data to have most impact.

**Characteristics of performance data**

Performance information will be most useful when it is defined and presented in such a way that it accurately and reliably measures the concepts being examined. This means that at times it will be necessary to explore new data sources rather than drawing on what is readily available.

Careful consideration should be given to deciding what type of measures to use (be these structural, outcome or process measures), taking into account the advantages and disadvantages of each. Outcome measures should usually be risk adjusted to account for variations in patient characteristics. While there has been substantial progress in developing risk adjustment tools, many methodological challenges remain. Moreover, the use of composite indicators has been widespread to obtain a broader perspective on performance. However, if they are not well designed they may be misleading.<sup>9</sup> Notwithstanding such causes for caution, ensuring that indicators are developed in a robust way, and applied properly will lead to performance measures that exhibit the characteristics of acceptability, feasibility, reliability, sensitivity to change and validity<sup>10</sup> (Box 1).

## Key challenges and future directions

The effectiveness of performance measurement is determined by the degree to which it assists in achieving a health system's objectives. Thus, while it is important to take into account statistical properties and methodologies, broader considerations of dissemination and the political and organizational context of performance measurement are also critical. Today, there is a great deal of political consensus on the need for performance measurement, yet modern health care systems are at different stages in development and implementation. It is clear, however, that international experience and experimentation can provide some lessons to aid future efforts:

1. A clear conceptual framework and a clear vision of the purpose of the performance measurement system should be developed. This framework should be aligned with accountability relationships inherent in the health system as well as other health system mechanisms (for example, finance, market structure, IT).
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 fit into a coherent conceptual framework.
4. Indicators should aim to measure data that is relevant to the needs of specific actors, and not focus merely on measuring what is available or easy to measure. More research is needed on what elements of performance are valued by patients, payers, purchasers and regulators.
5. Indicators should aim to be statistically sound, and be presented in a way that is straightforward to interpret, thus reducing the likelihood of susceptibility to manipulation or misinterpretation.
6. Indicators should be presented with full acknowledgement of any data limitations, including uncertainty estimates and lack of timeliness. Further exploration of improved processes for handling measurement errors is needed as such errors may confound true performance differences.
7. Attention should be given to enhancing capacity to understand and use information among managers and clinicians, making performance data an intrinsic part of clinical education and lifelong professional development.
8. Incentives to act upon 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.
9. More attention should be paid to the presentation of performance data and how this influences its interpretation by patients, providers and practitioners.
10. Policy makers should pay particular attention to the political and 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 process of performance management should be managed effec-

## Box 1: Qualities of good performance measures

### Development of indicators

**Face/content validity:** the extent to which the indicator accurately measures what it purports to measure.

**Reproducibility:** the extent to which the indicator would be the same if the method by which it was produced was repeated.

### Application of indicators

**Acceptability:** the extent to which the indicator is acceptable to those being assessed and those undertaking the assessment.

**Feasibility:** the extent to which valid, reliable and consistent data is available for collection.

**Reliability:** the extent to which there is minimal measurement error; the extent to which findings are reproducible should they be collected again by another organization.

**Sensitivity to change:** the extent to which the indicator has the capacity to detect changes in the unit of measurement.

**Predictive validity:** the extent to which the indicator has the ability to accurately predict.

Source: Campbell 2002<sup>10</sup>

tively. Amongst other things, this involves ensuring that specific interest groups do not capture the performance information system, and encouraging healthy political debate.

Health systems are still in the early days of performance measurement, and major steps are needed to improve data collection, methodologies, policy development and implementation. However, if properly applied, performance measurement offers immense potential for improving the accountability, responsiveness, equity and efficiency of health systems.

## REFERENCES

1. McIntyre D, Rogers L, Heiei EJ. Overview, history, and objectives of performance measurement. *Health Care Financing Review* 2001;22:7-43.
2. Smith PC. Performance measurement in health care: history, challenges and prospects, *Public Money & Management* 2005;25(4): 213-20.

3. World Health Organization (WHO) *The World Health Report 2000: Health Systems: Improving Performance*, Geneva: WHO, 2000.

4. Marshall MN, Shekelle PG, Davies HTO, Smith PC. Public reporting on quality in the United States and the United Kingdom, *Health Affairs* 2003;22(3):134–48

5. Dranove D, Kessler D, McClellan M, Statterthwaite M. Is more information better? The effects of 'report cards' on health care providers, *Journal of Political Economy* 2003;111(3): 555–88.

6. Dudley RA. Pay-for-Performance research: how to learn what clinicians and policy makers need to know, *JAMA* 2005;294(14): 1821–23.

7. Smith PC. On the unintended consequences of publishing performance data in the public sector, *International Journal of Public Administration* 1995;18(2–3):277–310.

8. Goodhart CAE. *Monetary Theory and Practice: The UK experience*. London: Macmillan, 1984.

9. Smith PC. Developing Composite Indicators for Assessing Health System Efficiency. In: Smith PC. *Measuring Up: Improving the Performance of Health Systems in OECD Countries*, Paris: OECD Publications, 2002.

10. Campbell SM, Braspenning J, Hutchinson A, Marshall M. Research methods used in developing and applying quality indicators in primary care, *Quality and Safety in Health Care* 2002;11:358–64.

*Irene Papanicolas is Research Associate in Health Economics at LSE Health, the London School of Economics & Political Science.*

*Peter C Smith is Professor of Economics and Director of the Centre for Health Economics at the University of York.*

*Elias Mossialos is Professor of Health Policy, and Director of LSE Health at the London School of Economics & Political Science and Co-Director of the European Observatory on Health Systems and Policies.*

## Long term care quality monitoring using the interRAI Common Clinical Assessment Language

Vincent Mor, Harriet Finne-Soveri, John Hirdes, Ruedi Gilgen and Jean-Noel DuPasquier

Residential long term care has been the mainstay of industrialized countries' long term care delivery systems for decades. However, changes in acute care financing, individuals' preferences for remaining in the community and the aging of the elderly population have all resulted in increasingly frail and impaired individuals occupying long term care facilities. As the acuity of long term care facilities has increased and as countries attempt to rebalance long term care investments by providing community support, the need to assess and monitor the quality of services provided to this frail and vulnerable population is paramount.

We describe the US designed nursing home Resident Assessment Instrument (RAI), or Minimum Data Set (MDS), and its adoption in several European countries' long term care systems. This minimum set of clinical and demographic data on all nursing home residents, originally intended as a clinical care planning tool, has been adapted in the US and beyond to determine payment levels and to monitor care quality. Similar applications are in place, either as a governmental mandate or on a voluntary basis, in Canada and several European countries, including Switzerland and Finland.<sup>1</sup>

### Origins

In the US, federal subsidy of long term care began once Medicare reimbursed for post-hospital nursing home and home care, and Medicaid began paying for nursing homes in 1966. Scandals about nursing home quality arose frequently, instigating investigations and commissions. In 1984, the Institute of Medicine

recommended various changes, most of which were translated into law passed in 1987, including a mandate to comprehensively assess all residents. Systematic assessment structures the clinical information necessary for care planning and provides the basis for a common lexicon.<sup>2</sup> An assessment was nationally implemented in 1991, updated in 1997 and universally computerized in 1998. Following considerable testing, the MDS was found to be reliable and generally valid in population-based research<sup>3,4</sup> and MDS data were correlated with research quality instruments for cognition, depression and physical function.<sup>5</sup>

The RAI was soon used for policy applications such as case-mix reimbursement which pays facilities differentially for serving more impaired and sicker patients. During the 1990s, many US states adopted prospective reimbursement based on case-mix, a trend that was accelerated by the universal availability of the MDS and the subsequent adoption of per diem reimbursement for Medicare-reimbursed skilled nursing facility stays.<sup>6</sup>

### Development of quality indicators

Creating quality indicators to monitor provider performance also became possible with the universal availability of the RAI, both to guide quality improvement efforts in a single nursing home and to generate and publicly report nursing home quality indicators. In 2002 the Centers for Medicare and Medicaid Services (CMS) began posting quality measures onto their "Nursing Home Compare" web site.<sup>7</sup> In spite of known

technical limitations of the measures, publicly reported data are now promulgated widely. A pay for performance demonstration project that rewards facilities based upon their quality performance and reductions in acute hospitalizations is now underway.<sup>8</sup>

### International application

Under the auspices of interRAI, an international organization of long term care researchers and policy advocates, the RAI has been modified for use in other countries' residential care facilities, home care services and even geriatric hospital settings. For example, there is widespread implementation of interRAI instruments in Canada and in Ontario chronic care hospitals participate in a score card reporting on clinical outcomes.\* While there was initial concern about publicly reporting results, transparency is now accepted. Other long term care facilities are scheduled to implement the required assessment system along with public reporting.

Consortia of long term care facilities in several European countries have adopted the RAI and disseminate quality reports to stimulate quality improvement efforts without public reporting. In Finland, the National Research and Development Center for Welfare and Health (STAKES) is responsible for identifying and promulgating "best practices" in elder care. RAI activities were launched by STAKES in 2000 as a pilot study, implementing the RAI assessment system in Finnish long-term care facilities. Facilities' participation is voluntary; however once committed, they assess all residents and contribute data to a central repository where reports summarizing facility case mix and quality measures comparing each facility with all others are generated. There are now over 350 facilities participating in the process, many involved in semi-annual training events focused on improving quality deficiencies identified in the reports.

In the late 1990s seven Cantons in the German-speaking part of Switzerland adopted the RAI for facility reimbursement and to measure quality. By 2006 there were over 300 facilities in seven Cantons serving over 20,000 residents contributing RAI data to a company called Q-Sys that generates semi-annual reports comparing each facility to peers within, as well as across, Cantons. The adoption of the RAI is attributable to the Health Insurance Law in 1996 which altered the basis of nursing home payment, resulting in a more uniform system of coverage for long term care in all Swiss Cantons.\*\* Regulations under the new law required that all residents of nursing homes wishing to be reimbursed undergo a geriatric assessment using a standardized instrument. Furthermore, nursing home providers were obligated to institute quality improvement programmes to be reimbursed; both conditions were met by the RAI.

### Summing up

Clinical assessment data can be used to create nursing home quality performance measures to identify areas for quality improvement. This is reinforced when providers come together as a consortium to share best practices in quality improvement strategies and track performance changes. Performance measures can also be used to assist quality assurance inspectors. Performance measures can be publicly reported to help consumers and their advocates select facilities and governmental entities can use them to adjust payment levels in accordance with quality. While there are still numerous conceptual and technical problems associated with interpreting differences among providers on the quality performance measures, the examples from the US, Canada, Finland and Switzerland clearly reveal that the impetus for quality improvement is greatly stimulated by comparative data.

### REFERENCES

1. Carpenter GI et al. Targeting and quality of nursing home care. A five-nation study. *Aging (Milano)* 1999;1(2):83-9.
2. Mor V. A comprehensive clinical assessment tool to inform policy and practice: applications of the minimum data set. *Medical Care* 2004;42(4 Suppl):III50-9.
3. Gambassi G et al. Validity of diagnostic and drug data in standardized nursing home resident assessments: potential for geriatric pharmacoepidemiology. SAGE Study Group. Systematic Assessment of Geriatric drug use via Epidemiology. *Medical Care* 1998;36(2):167-79.
4. Mor V et al. Inter-rater reliability of nursing home quality indicators in the US BMC. *Health Services Research* 2003;3(1):20.
5. Morris JN et al. MDS Cognitive Performance Scale. *Journal of Gerontology* 1994;49(4):M174-82.
6. Fries BE et al. Refining a case-mix measure for nursing homes: Resource Utilization Groups (RUG-III). *Medical Care* 1994;32(7):668-85.
7. Mor V. Improving the quality of long-term care with better information. *Milbank Quarterly* 2005;83(3):333-64.
8. Rahman A. The Pay for Performance demonstration program. *Journal of the American Medical Directors Association* 2006;7(2):133.

*Vincent Mor is Professor of Medical Science and Chair, Department of Community Health, Brown University School of Medicine, USA.*

*Harriet Finne-Soveri is Senior medical officer and RAI project manager at STAKES, Finland.*

*John P Hirdes is Professor in the Department of Health Studies and Gerontology, University of Waterloo, Canada.*

*Ruedi Gilgen is senior physician at Waid City Hospital Zürich, Geriatrics and rehabilitation, Switzerland.*

*Jean-Noël DuPasquier is CEO and senior economist and consultant at Me-Ti SA, Carouge, Switzerland.*

\* <http://www.cms.hhs.gov/NursingHomeQualityInits/Downloads/NHQISnapshot.pdf>; Accessed 18 February 2008.

\*\* <http://www.rai.ch/>; Accessed 27 December 2007.

# Public reporting of performance in the USA: The New York State CSRS

Paul Shekelle

The public reporting of performance data in health care continues to gain momentum worldwide. The country with the largest amount of experience in this area is the United States, and the New York State Cardiac Surgery Reporting System (CSRS) is one of the oldest and best known public reporting systems.

## The model

The CSRS had its genesis in 1989 when the Department of Health and its Cardiac Advisory Committee (a collection of nationally known experts) began an effort to reduce mortality following coronary artery bypass surgery (CABG). It began as a registry of clinical data collected on all patients undergoing cardiac surgery in New York State, with data about demographics, risk factors, and complications being collected prospectively under the supervision of the director of each cardiac surgery programme. These data were forwarded to the Department of Health, which developed a multivariate logistic regression risk adjustment model and used it to calculate expected and risk-adjusted in-hospital mortality rates by hospital and surgeon. Originally envisioned as a public reporting system of only hospital-level data, in response to a lawsuit filed by a newspaper the Department of Health was compelled to release surgeon-specific data, and later continued to do so for surgeons who performed at least 200 cases during a three-year reporting period and did at least one case in each year. Physicians and hospitals initially reacted very unfavourably to the public reporting of CSRS data, but seem to have adapted over time to accept the situation where public reporting of such data are considered routine.

Key to making comparisons in the CSRS

is the risk adjustment equation. About 40 risk factors are collected on each patient. In the most recent version of the CSRS (2002–2004) the model includes age and gender; hemodynamic status, left ventricular ejection fraction in four categories, history and timing of prior myocardial infarction, prior open heart surgery, and information about the co-morbidities of chronic obstructive airway disease and chronic kidney disease. The model, which continues to be refined, performs quite well at predicting mortality.

## Effects of public reporting

The effects of the CSRS have been studied and reported extensively in the literature. In a recent review of public reporting, we found 18 published studies assessing its effect on quality, making it by far the most studied public reporting system.<sup>1</sup> Despite this degree of study, controversy about the system's effects remains.

Cardiac surgery mortality rates have been trending downward for years throughout the developed world. The challenge has been attributing causality to the introduction of public reporting. Hannan et al.<sup>2</sup> reported a reduction in mortality (risk-adjusted mortality decreased from 4.17% to 2.45%) after the institution of the New York State CSRS. Likewise, in 1994 Dziuban et al.<sup>3</sup> found that risk-adjusted mortality improved from 6.6% in 1991 to 1.8% in 1993.

In the same year, another study<sup>4</sup> reported that while the risk-adjusted cardiac mortality of all surgeons and hospitals improved, it was noteworthy that providers with the highest initial mortalities displayed the most improvement. Moreover, Ghali et al.<sup>5</sup> found that observed mortality rates in Massachusetts (with no public reporting) decreased at a

rate comparable to those observed in New York and northern New England (regions that had publicly-released performance coronary artery bypass graft (CABG) data), questioning the effectiveness of publicly releasing performance results.

One of the concerns with public reporting by surgeons is that it may cause surgeons to decline to operate on more severely ill patients, because they believe this will adversely affect their publicly reported outcomes. Omoigui and colleagues<sup>6</sup> reported that the expected and observed mortality rates for New York residents who underwent CABG surgery at the Cleveland Clinic (outside New York state) increased after the 1991 release of the CSRS data, and postulated that there was selective referral out of state of sicker cardiac patients. Dranove and colleagues<sup>7</sup> reported that compared to states without public reporting, teaching hospitals in New York and in Pennsylvania (another state with public reporting of cardiac surgery) picked up an increasing share of patients requiring CABG surgery who also needed prolonged hospital stays, suggesting an increase in severity of illness.

The methodologically strongest study is that of Peterson and colleagues<sup>8</sup> who demonstrated that the reduction in mortality associated with cardiac surgery is greater in New York state than the national US trend (33% vs.19%), and found no evidence that there was decreased access to cardiac surgery among elderly patients with acute myocardial infarction or among higher-risk elderly subsets. This result, combined with the case studies reported by Chassin<sup>9</sup> showing high outlier hospitals took steps to improve their cardiac

surgery programmes, are sufficient for this analyst to conclude that there is a causal relationship between the CSRS and the reductions in 30-day mortality following CABG in New York State.

However, the gains from public reporting may lessen over time. One study<sup>10</sup> has shown that high-mortality New York hospitals had a reduction in mortality of 1.2% during the first 12 months after release of public reports, but the reductions were much less in subsequent years. No significant effect was found on low-mortality hospitals (i.e. those identified as good performers).

### Summing up

The New York State Cardiac Reporting System is a mature public reporting system that has probably had tangible benefits in terms of reducing the 30-day mortality rate following CABG surgery. Keys to its success have been the involvement of nationally known clinical experts in the design and implementation of the CSRS, the science behind the risk adjusting model, regular audits of the data, and being responsive to hospital and physician concerns about the accuracy and fairness of the data. These lessons should be helpful to other countries considering public reporting systems.

The CSRS is available at:

<http://www.health.state.ny.us/statistics/diseases/cardiovascular/>

### REFERENCES

1. Fung CH et al. Systematic review: the evidence that publishing patient care performance data improves quality of care, *Ann Intern Med* 2008;148(2):111–23.
2. Hannan EL et al. Improving the outcomes of coronary artery bypass surgery in New York State, *JAMA* 1994;271(10):761–66.
3. Dziuban SW Jr et al. How a New York cardiac surgery program uses outcomes data, *Ann Thorac Surg* 1994;58(6):1871–76.
4. Hannan EL et al. New York State's cardiac surgery reporting system: four years later, *Ann Thorac Surg* 1994;58(6):1852–57.
5. Ghali WA et al. Statewide quality improvement initiatives and mortality after cardiac surgery, *JAMA* 1997;277(5):379–82.
6. Omoigui NA et al. Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes, *Circulation* 1996;93(1):27–33.
7. Dranove DEA. Is more information better? The effects of "Report Cards" on health care providers. *Journal of Political Economy* 2003;111(3):555–88.
8. Peterson ED et al. The effects of New York's bypass surgery provider profiling on access to care and patient outcomes in the elderly, *J Am Coll Cardiol* 1998;32(4):993–99.
9. Chassin MR. Achieving and sustaining improved quality: lessons from New York State and cardiac surgery, *Health Affairs (Millwood)* 2002;21(4):40–51.
10. Cutler DM, Huckman RS, Landrum MB. The role of information in medical markets: an analysis of publicly reported outcomes in cardiac surgery. In: *National Bureau of Economic Research, Inc, NBER Working Papers*, 2004.

*Paul Shekelle is Director of the Southern California Evidence-Based Practice Center, RAND Corporation; staff physician at the West Los Angeles Veterans Affairs Medical Center; and Professor at the UCLA School of Medicine.*

## Using composite indicators to measure performance in health care

Maria Goddard and Rowena Jacobs

Much attention is given to the relative performance of health care systems, organizations, services and individual practitioners. The public and the policy maker have a legitimate interest in the many and varied aspects of performance such as quality, efficiency, responsiveness, equity and outcomes and there has been an explosion of interest in the generation, publication and interpretation of performance information in the health care domain across the world. However, the very abundance of such information can obscure the ability of consumers and policy makers to make overall judgments about relative performance. Complex information presented over a large number of dimensions may be difficult to comprehend and a lack of transparency presents opportunities for poor

performance to go undetected; hence, the popularity of composite indicators which are an aggregate of a number of underlying performance measures. They have instant appeal because they provide a single score or rating that is easy to understand and they offer an attractive way of summarizing a wealth of performance data, often facilitating the production of rankings or league tables

### Methodological challenges

Despite the widespread use of composite indicators, their construction presents many methodological challenges which, if not treated carefully and transparently, can leave them open to misinterpretation and potential manipulation. The accuracy, reliability and appropriateness of such indices need to be explored if major

policy, financial and social decisions hinge on the performance of organizations as measured by composite indicators.

It is important to consider some of the key steps in constructing and using a composite indicator, covering issues such as choosing the underlying indicators to be used; transforming the indicators; applying a system of weights to the indicators; and combining them all to form the new performance measure.

We highlight here just two of the issues that have the most impact on the usefulness of composite performance measures.

### Choice of indicators

First, the decision about which indicators to include in the composite is crucial. Careful judgement is required as the impact will be to focus effort on the included indicators, potentially at the expense of the excluded indicators. Composites are often either opportunistic and/or incomplete (measuring aspects of performance that are captured in existing data), or are based on highly questionable sources of data. Either weakness can cause serious damage to the credibility of the composite. The choice of indicators is most often constrained by data availability and may therefore give an unbalanced picture of health services. The excluded indicators may be as (or more) important than the included indicators but simply more difficult to measure.

The higher the level at which composites are created and the broader their scope, the greater will be the issues of data availability and lack of comparability. For example, the WHO composite index of health system performance was produced for 191 countries and sought to be comprehensive in coverage, measuring five domains: overall health outcomes; inequality in health; fairness of financing; overall health system responsiveness; and inequality in health system responsiveness.<sup>1</sup> It is clear that there will be a trade-off between an ambitious aim of deriving a composite measure capturing complex and comprehensive health performance dimensions for a wide range of countries,

and the practical issues of gathering good data on such dimensions.

At the other end of the scale, composite measures for more homogenous organizations such as hospitals can focus on a smaller range of measures, chosen either to reflect health priorities (for example, those with the largest impact on health) or because they are most amenable to managerial intervention and improvement. Even where there has been criticism of the quality and comparability of the data underlying composites, it is often the case that merely using the data provides an incentive to improve accuracy for future performance measurement exercises.

### Weighting and decision rules

The second issue that affects substantially the reliability and robustness of composite measures is the methodology used to combine the underlying indicators into a single score. All variables may be given equal weight or they may be given different weights which reflect the priority, reliability or other characteristics of the underlying measures. Weights are essentially value judgements about the importance of different performance indicators and about the relative opportunity cost of achieving those performance measures. The incentive effect of weighting is potentially very powerful and weighting schemes can change dramatically the ranking of a particular organization if an indicator is given more weight on which the organization either excels or fails.

There are many alternative methods of creating weights, ranging from the opinion of experts or the general public to the use of complex statistical techniques. The key point to emerge from considering the experience of alternative weighting systems in the health sector is that most weighting systems will contain an element of subjectivity. Therefore, we should remain alert to this when inter-

preting results and look for good practice which explores the sensitivity of results to alternative sets of weights.

Further complexity is introduced where the composites are combined using decision rules which can sometimes lack transparency.\* Research suggests that performance ratings are very sensitive to small changes in decision rules although there are often good reasons for using such an approach. For example, they can ensure that certain minimum requirements are met by only allowing a good score if organizations do well on individual indicators that are felt to be most important.

### Summing up

At each stage in the construction of composite indicators, choices are made that may appear to be largely technical or may be thought to have minor significance, but can in fact have a fundamental impact on the final results. Whilst this may call into question the utility of composite scores and suggest that an array of performance data may be more appropriate, we argue that composite scores have an important role to play in helping to focus attention on key aspects of performance in a way that the public and policy maker can understand.

### REFERENCES

1. World Health Organization. *The World Health Report 2000, Health Systems: Improving Performance*, WHO: Geneva, 2000.

---

*Maria Goddard is Professor of Health Economics at the Centre for Health Economics, University of York, UK.*

*Rowena Jacobs is Senior Research Fellow at the Centre for Health Economics, University of York, UK.*

---

\* Decision rules reflect views on the importance of achieving certain standards. They set the boundaries within which performance scores will be allocated (for example, defining what constitutes a “good” or “poor” score on an indicator); or they may disallow an organization from obtaining a good performance score if they fail to meet a target. Such rules are often applied sequentially and implicitly introduce a set of weights.

# Attribution and causality bias in health care performance measurement

Darcey D Terris and David C Aron

A desirable health care performance measure accurately reflects the quality of care provided by the source of care under analysis. When quality measures are biased, accountability is unfairly directed and improvement is inappropriately targeted. Accurate assessment of health care quality is dependent on fairly dividing (attributing) and assigning responsibility (causality) for the observed processes and outcomes of care among many factors, including the source of care.

To clarify the challenge presented, let's consider an episode where it is observed that a patient did not receive a guideline-recommended treatment. Assuming receipt of treatment is an indicator of quality, this episode points to a possible lapse by the attending physician. The doctor may have neglected to prescribe the treatment, as appropriate, given the patient's diagnosis and co-morbidity. Alternatively, the doctor may have prescribed the treatment, with the patient refusing the recommendation due to their health beliefs. Perhaps the patient was willing, but access was restricted by their insurance carrier? Limited accessibility of facilities capable of dispensing the treatment may have created an insurmountable barrier. Finally, the patient may have received the treatment, but information systems in place failed to record the event.

## Risk assessment and confounders

How do we, in our quality assessments, account for these many alternative explanations? Typically, statistical modelling is used to evaluate the relationship between a given source of care and a process or outcome variable identified as a quality indicator. Through a process of 'risk adjustment' control variables are included

in the model to account for additional factors (called 'confounders') that may influence the incidence of the quality indicator. Adequate risk adjustment is limited by our knowledge and acknowledgement of potential confounders and our ability, and available resources, to measure confounders for inclusion in quality assessments. In the previous example, confounding was hypothesized to arise from various levels, including patient characteristics (health beliefs), practice resources (information systems), health system policies (coverage), and the community where the patient resides (accessible facilities).

Similar hurdles may limit a physician's control over care recommendations. For example, health system policy may restrict the number of referrals within a given period. Non-emergent patients who present after the referral limit is reached may be asked to return at a later date to receive the referral. Depending on timing, a performance assessment may indicate that the recommended process of care did not occur for these patients. Thus, even if a causal relationship has been established, responsibility for the observed process of care within a given context may not be attributable to the source of care, but may be directed by local, regional and national health system and regulatory bodies.<sup>1,2</sup>

## Greatest risk of bias

Which sources of care are at greatest risk for bias in health care performance assessment? Health care providers in resource-limited settings are at greater risk, as compared to their counterparts in more resource-affluent settings.<sup>3,4</sup> In part, this disparity arises from differences in the locus of control for acquiring and

directing the use of health care resources.<sup>5</sup> Community resources (for example, neighbourhood socioeconomic status, local public health practice, general infrastructure, etc.) also influence the health, and health care processes and outcomes, obtained by a community's residents.<sup>4</sup>

Outside of setting, a source of care that treats more complex patients is also at greater risk of bias in health care performance measurement.<sup>4</sup> This complexity arises from the health status of the patient (for example, co-morbidity) and other patient-level characteristics (for example, health beliefs and behaviour). More complex patients are typically found in resource-limited settings<sup>3</sup>, within teaching hospitals<sup>6</sup>, and in settings that specialize in complex patient subgroups, such as the elderly or disabled.<sup>7</sup>

## Impact

If bias in health care performance assessment occurs, the signal that a health care policy or practice should be repeated (or avoided) is lost. High quality doctors may not sustain their current practice policies and may substitute new initiatives that reduce the quality of care provided. Conversely, low quality doctors, inaccurately assessed as providing higher quality care, will not receive the clear signal needed to target service delivery improvements. When performance measurement is linked to reimbursement or other market-based incentives, perceptions of the risks associated with inaccurate assessment may create disincentives contrary to the goal of improving health care quality and equity.<sup>8</sup> Under these circumstances, physicians may avoid including complex patients in their practice or locating their practice in more

complex settings, creating difficulties in recruitment and retention of doctors for disadvantaged population segments.<sup>3</sup>

### Policy implications

The risk of causality and attribution bias can be minimized by following these key steps in development and implementation of health care performance measurement policies:

1. Systematically review existing reports investigating the possibility of a causal and attributable link between the source of care under assessment and the process or outcome of care proposed as a quality indicator.
2. Perform a prospective analysis to identify the critical pathways involved in achieving desired and undesired processes and outcomes of care. Identify possible confounders to the relationship between the source of care under assessment and the proposed quality indicator.
3. Synthesize the results of steps 1 and 2 and identify essential gaps in knowledge. Involve stakeholders, both internal and external to the source of care under analysis, in the process. If a new study is required, consider sources of random and systematic error in measurement and sampling when developing the study design.
4. Employ risk adjustment and advanced modelling techniques to account for confounding from multiple levels of the health care environment when investigating a possible causal and attributable link between a source of care under assessment and the proposed quality indicator.
5. Acknowledge that causality and attribution bias cannot be completely eliminated. Consider unintended impacts from experienced or perceived bias in quality assessment on

future improvement of health care quality and equity.

### REFERENCES

1. Landon BE, Reschovsky J, Reed M, Blumenthal D. Personal, organizational, and market level influences on physicians' practice patterns: results of a national survey of primary care physicians, *Medical Care* 2001;39(8):889–905.
2. Hauk K, Rice N, Smith P. The influence of health care organizations on health system performance, *Journal of Health System Research & Policy* 2003;8(2):68–74.
3. Casalino L, Elster A. Will pay-for-performance and quality reporting affect health disparities? *Health Affairs* 2007;26(3):w405–w414.
4. Terris DD, Litaker DG. Data quality bias: an under-recognized source of misclassification in pay for performance reporting? *Quality Management in Health Care* 2008;17(1):19–26.
5. Miller R, West C. The value of electronic health records in community health centers: policy implications, *Health Affairs* 2007;26(1):206–14.
6. Antioch KM, Ellis RP, Gillett S, Borovnicar D, Marshall RP. Risk adjustment policy options for casemix funding: international lessons in financing reform, *European Journal of Health Economics* 2007;8(3):195–212.
7. Jette AM, Smith KW, McDermott SM. Quality of medicare-reimbursed home health care, *Gerontologist* 1996;36(4):492–501.
8. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma, *The Lancet* 2004;363(9424):1147–54.

Darcey D Terris is a Senior Scientist at the Mannheim Institute of Public Health, Social and Preventive Medicine, Mannheim Medical Faculty, University of Heidelberg, Mannheim, Germany.

David C Aron is Director of the VA HSR&D Center for Quality Improvement Research, Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, Ohio, USA.

## New Health Systems in Transition (HiT) profiles

These latest additions to the HIT series are also available online.



**Denmark** is a small country with 5.4 million inhabitants. Danish health care is dominated by the public sector, financed by local and state taxes and administered by

the regions. In recent years the focus of health care reform has been on patient choice, waiting times, quality assurance and coordination of care.

The publication of the HiT is very timely because a major structural reform in 2007 changed the political and administrative landscape dramatically. Reforms in the way health care is financed also took place.

Available at <http://www.euro.who.int/Document/E91190.pdf>



**Kazakhstan** – this HiT allows the reader an insight into a troubled society and one which, in the 1990s, was beset by economic problems and uncertainty as to

how to undertake essential reforms. This led to a period of stagnation and a serious rise in mortality rates. In the wake of an economic upswing fuelled by oil revenues in recent years, Kazakhstan found itself in an improved position. Using some of its new resources it has embarked on a comprehensive national health reform programme which is aimed to transform the situation by 2010.

Available at <http://www.euro.who.int/Document/E90977.pdf>



World Health Organization  
Regional Office for Europe



Government of Belgium



Government of Finland



Government of Greece



Government of Norway



Government of Spain



Government of Sweden



European Investment Bank



Veneto Region of Italy



Open Society Institute



World Bank



London School of Economics and Political Science



London School of Hygiene & Tropical Medicine

## FORTHCOMING OBSERVATORY STUDY

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

#### Edited by

**Peter C Smith (Centre for Health Economics, University of York)**

**Elias Mossialos (European Observatory on Health Systems and Policies and London School of Economics & Political Science)**

**Sheila Leatherman (School of Public Health, University of North Carolina)**

**Irene Papanicolaos (LSE Health, London School of Economics & Political Science)**

Information is crucial to the delivery of efficient and effective health care. Not only does it serve as a fundamental resource for securing appropriate treatment and good outcomes for patients, it is a key resource for securing managerial, political and democratic control of the health system.

The last 25 years has seen astonishing developments in the scope, nature and timeliness of performance data available in all developed health systems. Yet, there remain many unresolved discussions on the collection of performance data, how to analyse and report it, and on its validity and usefulness as part of an integrated framework of regulation, accountability and health system improvement.

This book summarizes some of the principal themes emerging in the performance measurement debate, examining experience to date, and offering guidance on future research and policy priorities. It provides a comprehensive policy discussion covering the opportunities and challenges associated with performance measurement; the various dimensions and levels of health system performance; the analytical tools needed to implement successful performance measurement; the design and implementation of performance measurement systems; and the implications of performance measurement for policy makers, politicians, regulators, and others charged with the governance of health systems.

Published by Cambridge University Press in 2009.

## Contents

### 1. Principles of performance measurement (Editors)

#### 2. The dimensions of performance

Population health (Chris Bain, University of Queensland, Martin McKee and Ellen Nolte, The London School of Hygiene & Tropical Medicine)

Individual health outcomes from health care interventions (Ray Fitzpatrick, University of Oxford)

Measuring clinical quality and appropriateness (Elizabeth McGlynn, RAND, Alessandro Liberati, University of Modena)

Financial protection (Adam Wagstaff, World Bank)

Responsiveness (Somnath Chatterji, Amit Prasad and Nicole Valentine (WHO) and Nigel Rice and Silvana Robone, University of York)

Measuring equity in the health system (Xander Koolman, Erasmus University)

Health system productivity (Unto Hakkinen, STAKES and Andrew Street, University of York)

#### 3. Analytical methodology for performance measurement

Risk adjustment in performance measurement (Lisa Iezzoni, Harvard University)

Clinical surveillance and patient safety (Olivia Grigg and David Spiegelhalter, University of Cambridge)

Attribution and causality (Darcey Terris, Heidelberg University, David Aron, US Veterans' Administration)

Composite measures of performance (Maria Goddard and Rowena Jacobs, University of York)

#### 4. Performance measurement in specific domains

Primary care (Helen Lester and Martin Roland, University of Manchester)

Chronic care (Martin McKee and Ellen Nolte, The London School of Hygiene & Tropical Medicine)

Mental illness (Rowena Jacobs, University of York and David McDaid, The London School of Economics & Political Science)

Long term care (Vincent Mor, Brown University, Harriet Finne-Soveri, STAKES, John P Hirdes, University of Waterloo, Ruedi Gilgen, Waid City Hospital Zurich, Jean-Noel DuPasquier, Me-Ti SA, Switzerland)

#### 5. Health policy and performance measurement

Targets and performance measurement (Reinhard Busse, Berlin University of Technology)

Public performance reporting (Paul Shekelle, RAND and UCLA)

Information technology and performance measurement (Tom Sequist, and David Bates, Harvard University)

Incentives for promoting better performance (Doug Conrad, University of Washington)

Performance information for professional improvement (Arnold Epstein, Harvard University)

International Comparison (Jeremy Veillard, WHO Regional Office for Europe, Sandra Garcia, OECD, Sowmya Kadandale and Niek Klazinga, University of Amsterdam)

#### 6. Conclusions (Editors)

## Editor

Anna Maresso

## Editorial Team

Josep Figueras

Martin McKee

Elias Mossialos

Ellen Nolte

Sarah Thomson

## To join the mailing list, please contact

Anna Maresso

Observatory – London Hub

Tel: +44 20 7955 6288

Fax: +44 20 7955 6803

Email: [a.maresso@lse.ac.uk](mailto:a.maresso@lse.ac.uk)

*Euro Observer* is published quarterly by the European Observatory on Health Systems and Policies.

The views expressed in *Euro Observer* are those of the authors alone and not necessarily those of the European Observatory on Health Systems and Policies or its participating organizations.

© European Observatory on Health Systems and Policies 2008.

No part of this document may be copied, reproduced, stored in a retrieval system or transmitted in any form without the express written consent of the European Observatory on Health Systems and Policies.

For information and ordering details on any of the Observatory publications mentioned in this issue, please contact:

The European Observatory on Health Systems and Policies  
WHO ECHP

Rue de l'Autonomie, 4

B - 1070 Brussels, Belgium

Tel: +32 2 525 09 33

Fax: +32 2 525 0936

Email: [info@obs.euro.who.int](mailto:info@obs.euro.who.int)

Website: [www.observatory.dk](http://www.observatory.dk)

## Design and production by

Westminster European

[westminster.european@btinternet.com](mailto:westminster.european@btinternet.com)