

LSE

THE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE ■

LSE Research Online

Peter C. Smith, [Irene Papanicolas](#)

Health system performance comparison: an agenda for policy, information and research

Policy summary

Original citation:

Papanicolas, Irene and Smith, Peter, eds. (2013) *Health system performance comparison: an agenda for policy, information and research*. Policy summary , 4 . [Open University Press](#), Maidenhead, UK. ISBN 9780335247264

Originally available from [European Observatory on Health Systems and Policies](#)

This version available at: <http://eprints.lse.ac.uk/54802/>

Available in LSE Research Online: December 2013

© 2012 World Health Organization and World Health Organization, on behalf of the European Observatory on Health Systems and Policies

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.



World Health
Organization

REGIONAL OFFICE FOR **Europe**

POLICY SUMMARY 4

Health system performance comparison: an agenda for policy, information and research

Peter C. Smith, Irene Papanicolas

Keywords:

**HEALTH SYSTEMS PLANS –
organization and
administration**

**DELIVERY OF HEALTH CARE –
organization and
administration**

**OUTCOME ASSESSMENT
(HEALTH CARE)**

COMPARATIVE ANALYSIS

HEALTH CARE SURVEYS

HEALTH POLICY

EUROPE

This policy summary is one of a new series to meet the needs of policy-makers and health system managers.

The aim is to develop key messages to support evidence-informed policymaking, and the editors will continue to strengthen the series by working with authors to improve the consideration given to policy options and implementation.

© World Health Organization 2012 and World Health Organization, on behalf of the European Observatory on Health Systems and Policies 2012

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

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

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

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

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

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

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

POLICY SUMMARY 4

Health system performance comparison: an agenda for policy, information and research

Peter C Smith, Irene Papanicolas

This work is a summary of material prepared for a forthcoming book by the European Observatory on Health Systems and Policies. The contributions of all authors to the book and this summary are gratefully acknowledged.

Health system performance comparison: an agenda for policy, information and research

Contents	Page
Key messages	iv
1 Policy issue: why use international comparisons?	1
2 Lessons from international comparisons to date	2
3 How to compare key domains of performance	8
4 Future directions and conclusions	32
References	37

Authors

Irene Papanicolas, Lecturer in Health Economics, Department of Social Policy, London School of Economics and Political Science, London, United Kingdom

Peter C. Smith, Professor of Health Policy, Imperial College Business School and the Centre for Health Policy, London, United Kingdom

Editors

WHO Regional Office for Europe and European Observatory on Health Systems and Policies

Editor

Govin Permanand

Editorial Board

Josep Figueras

Claudia Stein

John Lavis

David McDaid

Elias Mossialos

Managing Editors

Kate Willows Frantzen

Jonathan North

Caroline White

The authors and editors are grateful to the reviewers who commented on this publication and contributed their expertise.

No: 4

ISSN 2077-1576

Key messages

Policy issue

- International health system performance comparisons have the potential to provide a rich source of evidence as well as policy influence.
- Country comparisons that are not conducted with properly validated measures and unbiased policy interpretations may prompt adverse policy impacts and so caution is required in the selection of indicators, the methodologies used, and the interpretations made.

Lessons from international comparisons

- International health system comparisons provided by multilateral institutions such as the WHO or the OECD have generated much interest since the publication of the 2000 *World Health Report*, which highlighted the potential for cross-country learning from the scrutiny of comparable data, and also the challenges that must be overcome:
 - Definitions of performance indicators should be clear and consistent, and fit into a clear conceptual framework.
 - The metrics used in international comparison should enjoy widespread acceptance, and are defined in unambiguous terms that are consistent with most countries' data collection systems.
 - In order to draw meaningful comparisons and understand the drivers of differences in measures between systems it is usually necessary to adjust for variations in the demographic, social, cultural and economic circumstances of nations.
 - "Single number" measures of whole health system performance, while offering a more rounded view of performance, have limited scope for policy action, and may distract policy-makers from seeking out and remedying the parts of their system that require attention.
- Lessons from benchmarking activities in other sectors suggest that – when applied to health systems – benchmarking will be most effective if it focuses on practice as well as performance; is grounded in the broader change process; is well structured and planned in order to engage stakeholders; and carefully considers how performance is linked to resource allocation.

Lessons from recent experience in health system comparison

- In defining the boundaries of the health system it is important to be aware of the benefits of choosing both narrow and wide boundaries. Narrow boundaries are better suited to holding stakeholders accountable, while broader boundaries are better for a more holistic understanding of the determinants of valued outcomes.
- Performance measurement evaluates the extent to which a health system meets its key objectives. However progress in the development of data collection techniques in the different dimensions of health performance is variable.
- While efforts should be undertaken to improve data collection efforts, policy-makers should also make themselves familiar with limitations in existing indicators in order to be able to interpret them appropriately.

Comparing key domains of performance

- Population health measures often take a broad perspective, which captures the effect of many determinants of health beyond the delivery of health care. This broad perspective can be attractive from a political point of view because it draws attention to the importance of many sectors in determining health outcomes. However it also creates major methodological challenges in seeking to attribute changes in health to any particular actions. A narrower perspective, such as the concept of avoidable mortality, focuses on measures that can more easily be attributed to health care, and which are therefore more amenable to immediate health policy.
- All population health indicators suffer from a number of methodological problems, which need to be addressed in order to make international comparisons more meaningful. Some of the main issues involve availability and coding of data, particularly of data on cause of death where there are problems of comparability among countries and over time.
- Direct indicators of the contribution of health services to health status are available in the form of health service quality measures, such as standardized hospital mortality rates and numerous disease-specific health outcome measures, such as mortality rates, adverse events and complications. Far less prevalent are broader outcome measures in areas such as disabilities and discomfort.
- While existing health service measures offer some indicators of the performance of individual organizations, international comparison is complicated by different organizational settings and reporting

conventions, even after suitable adjustment for case-mix and other contextual circumstances.

- Although comparative indicators on inequality of health and equity in access to health care are available at both European and non-European level, equity indicators derived from existing projects and datasets may be misleading for policy-makers due to limitations in the availability and comparability of data.
- The many aspects of financial protection have yielded some important indicators, such as the incidence of catastrophic expenditure. However it has proved difficult to develop a single indicator capturing the full extent to which people are financially protected from health shocks. Currently, measures of the incidence and magnitude of households' direct payments for health care form the basis of metrics for financial protection assessment and system comparisons.
- There is still lack of clarity as to what dimensions should be included in the domain of responsiveness, which embraces concepts such as respect, confidentiality and prompt attention. This uncertainty leads to the measurement of different areas using different tools with different weights, domains and indicators that are difficult to summarize and compare.
- Efficiency indicators serve as a summary measure of the extent to which the inputs to the health system, in the form of expenditures and other resources, are used wisely to secure the goals of the health system. Almost all efficiency indicators are constructed as a ratio of inputs to outputs, offering an indication of the extent to which resources have been wasted along some or all of the production pathway.
- In measuring efficiency, a fundamental challenge is the assignment of inputs and associated costs to specific health system activities, often relying on arbitrary accounting rules or other questionable assignments. In principle, the inputs used should be directly and fully aligned with the output under scrutiny.

Future directions

- The key requirements for creating comparable indicators that address the needs of policy-makers are: appropriate methods of summarizing complex information; a narrative that picks out the key issues and uncertainties; a diagnosis of why the reported variations are arising; and an assessment of the implications for policy action.

1 Policy issue: why use international comparisons?

Individual nations are increasingly seeking to introduce more systematic ways of assessing the performance of their health systems, and of benchmarking performance against other countries. Policy-makers recognize that without measurement and comparison, it is difficult to identify good and bad delivery practice or good or bad practitioners (“what or who works”), to design health system reforms, to protect patients and payers, or to make the case for investing in health care. Measurement is also central to promoting accountability to citizens, patients and payers for health system actions and outcomes. This focus on assessment coincides with the enormous increase in the capacity for measurement and analysis of the last decade, driven in no small part by massive changes in information technology and associated advances in measurement methodology.

However, notwithstanding major progress by organizations such as the European Commission, the OECD, the Commonwealth Fund, the WHO and individual countries, performance comparison efforts are in their early stages, and there are many challenges involved in the design and implementation of comparison schemes.

The state of current developments in performance measurement is comprehensively surveyed in the book that followed the Tallinn conference: *Performance measurement for health system improvement* (Smith et al., 2009). The book identified the important sources of international comparison noted above, but also highlighted the limits in many performance assessment initiatives both in terms of their scope and their policy usefulness. It exposed the difficulties of interpreting sources of performance information from a health system policy perspective. It also draws attention to the danger that the existing focus on partial aspects of performance may lead to serious policy misconceptions if not accompanied by a careful commentary on the implications of variations for health system improvement and reform. Care and review are all the more important in light of the growing appetite for cross-country performance comparisons and benchmarking by Member States, citizens and the media, and the pressures that can arise from it.

In sum, properly conducted country comparisons of performance may constitute a rich source of evidence and powerful influence on policy. However, caution is required as initiatives that rely on poorly validated measures and biased policy interpretations may lead to ineffective or even adverse policy choices.

There is therefore a need to harness the potential of comparative health systems performance assessment (HSPA), building on credible initiatives and strengthening both the methodologies and the policy analysis. This includes highlighting not only the policy “uses” but also the policy “abuses” of

comparisons. In other words, as well as drawing out the information content and potential of performance measures, analysts should indicate what cannot be inferred from the analysis, showing the limitations of current measures and suggesting fruitful future improvements.

This policy brief seeks to summarize the current “state of the art” of health system comparison, identifying data and methodological issues and exploring the current interface between evidence and practice. It also draws out the priorities for future work on performance comparison, in the development of measurement instruments, analytic methodology, and assessment of evidence on performance. It will conclude by presenting key lessons and future priorities policy-makers should take into account.

2 Lessons from international comparisons to date

a. Lessons from international comparisons in health

International health system comparisons provided by multilateral institutions such as the WHO or the OECD have generated much interest since the publication of the 2000 *World health report* (WHO, 2000). While desultory performance measurement efforts occurred long before the report, dating back over a century, this publication highlighted the potential for cross-country learning from the scrutiny of comparable data. Through careful examination of the lessons from the 2000 *World health report* and other large-scale international health system comparisons that have been developed since, some conclusions can be drawn regarding lessons that this practical experience can offer.

There has been a fierce debate over the virtues of producing a “single number” measure of whole health system performance. Such composite measures offer a more rounded view of performance than a series of fragmentary metrics that offer insights into the performance of parts of the health system. However, they may disguise weaknesses in specific parts of health systems, and are difficult to interpret. Experience suggests that at the current stage of measurement and methodological development the usefulness of “whole-system” performance comparison is open to question, especially when used for global rankings of health systems. Such endeavours are readily open to challenge and may distract policy-makers from identifying and remedying the parts of their system that require attention.

Most performance measures are contestable and exhibit shortcomings of some sort. This calls for careful commentaries on existing data and better understanding of the reasons for variations between health systems. From an international comparative perspective, the crucial requirement is that such metrics are relevant to health system goals, enjoy widespread acceptance, and

are defined in unambiguous terms that are consistent with most countries' data collection systems. Where there may be discrepancies in data collection or reporting this should be presented clearly alongside the metric.

Progress has been made in certain areas, such as some aspects of health system quality comparison and accounting conventions. However, there are many domains of health system performance where there is no consensus on how metrics are to be conceptualized or measured, such as responsiveness, equity and efficiency. Moreover, in order to draw meaningful comparisons and understand the drivers of differences in measures between systems it is usually necessary to adjust for variations in the demographic, social, cultural and economic circumstances of nations. Many analyses make rudimentary adjustments for variation in demographic profiles, but more advanced progress in this area has been very limited, partly because a lack of information available to adjust for such circumstances. There is therefore clearly a major international agenda to be addressed to agree a conceptual framework for collecting comparative information, the domains that require measures, and the specification of those measures.

For assessment purposes, moreover, particular care should be taken when examining just a single snapshot (cross-section) of comparative performance. The use of time series can offer more secure inferences, but places greater demands on data availability. Outcomes (such as mortality) are often the product of the inputs of previous years, and will not necessarily be a reflection of the performance of the current health system. Conversely, current inputs may contribute in part to future attainment. A naive comparison of current levels of attainment might ignore the *trajectory* of the health system, for example attributing good current performance to current efforts, rather than (say) preventative programmes in the past. It is therefore highly desirable that any comparison of such outcomes takes into account these time lags – the comparison should be dynamic.

b. Lessons from other sectors

When considering how international comparisons can be most beneficial in health systems it is important to look beyond what has been done in health and learn from the experience of benchmarking efforts in other sectors. Indeed there is a widespread interest in and take up of benchmarking that appears to be an international phenomenon. Different forms of benchmarking have different aims and different lessons. A key distinction can be made between performance benchmarking and practice benchmarking. Performance benchmarking concentrates on establishing performance standards, while practice benchmarking is concerned with establishing the reasons why organizations achieve the level of performance that they do. Evidence suggests that performance benchmarking

is more prevalent than practice benchmarking, although it can be argued that practice benchmarking is more beneficial in the long run.

There are practical challenges to benchmarking. Deciding what and how to benchmark, which organizations to compare with and how to ensure comparability, what data to use and whether the data are robust, all affect its value. Particularly in practice benchmarking, one does not always need to seek organizational comparability. Indeed, when focusing on how to improve processes within organizations it is often helpful to compare organizations from very different sectors. One of the original proponents of benchmarking, Xerox, compared itself with LL Bean, a mail order company because Xerox wanted to improve its warehousing and distribution processes and LL Bean was recognized as having excellent warehousing and distribution processes.

Benchmarking is at its most impactful when integrated with a broader programme of change. This raises questions of who should benchmark (self-assessment versus external agencies) and to whom the subsequent information should be released. If benchmarking is carried out by external agencies and the data subsequently made public (as is often the case in the health sector) then there is a strong incentive for those being benchmarked to become defensive and seek to paint their organization in the best possible light. In extreme cases, this can result in perverse and dysfunctional behaviour, where people take actions that make the benchmark numbers look better, even though these actions do not really improve organizational performance.

The most common reasons for benchmarking are to improve efficiency and reduce cost, although the benefits of benchmarking are not limited to efficiency gains. Indeed benchmarking can deliver new insights that improve the quality of service delivered. Given the current financial situation in many countries, however, it is likely that we will see a strong focus on benchmarking for efficiency gains – identifying ways in which we can do more for less (or perhaps less for less).

In addition to the direct benefits of benchmarking, there can be indirect benefits. These apply especially in terms of cultural change and opening up the organization to new ideas. Encouraging staff to visit and understand the way other organizations work can be very powerful in opening staff's eyes. Indeed it is claimed that the Toyota Production System was inspired partly by visits from Taiichi Ohno and colleagues to American supermarkets, where they observed a much smoother flow of work than one would traditionally find in manufacturing plants.

Clearly there are challenges of benchmarking, as well as benefits. Indeed, some organizations claim to have achieved no benefit from their benchmarking activities, while others claim to have achieved significant benefit. An important

distinction between these two groups – those that claim to benefit from benchmarking and those that do not – is the extent to which benchmarking focuses on strategic priorities, involves clear and careful planning, and any practices identified are not simply adopted wholesale, but instead adapted appropriately for the adopting organization.

Questions remain about whether benchmarks should be linked to incentives and/or budgets. Clearly, for reasons of accountability and transparency, this might be useful. Additionally, in the not-for-profit sector there is some attempt to use benchmarks as a replacement for the market mechanism. Given there are no formal competitors for many public services, benchmarks provide a form of competitive comparison. Box 1 illustrates some of the lessons that can be applied from other areas to benchmarking in health systems.

Box 1. Lessons of benchmarking for health systems

From what is known of benchmarking in other sectors, five implications can be extended to benchmarking efforts in the health system:

1. Health benchmarks should focus on practice as well as performance.
2. Health benchmarks should not be used simply to evaluate and compare performance.
3. Benchmarks need to be grounded in a broader change process.
4. The benchmarking process itself needs to be well structured, well planned and designed to engage people in making change in their organizations.
5. The designers of health benchmarking systems need to consider very carefully the link between resource allocation and benchmark performance if they are to avoid dysfunctional behaviour.

Source: Adapted from Neely, forthcoming.

c. Frameworks for comparison

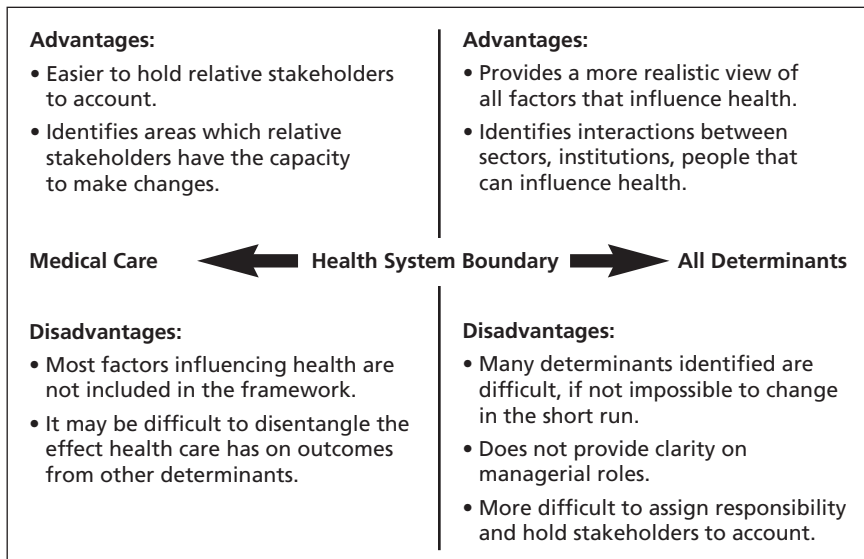
International comparison of health system performance can exert a major influence on national policy-makers. The response to the *World health report 2000* was an indication of the potential power of such comparison, but also highlighted the severe methodological difficulties that arise when seeking to make it operational. The starting point of most international comparisons is the creation of a conceptual framework on which to base the collection of information and to use as a heuristic for the understanding of the health system.

In the past decade much energy has been put into the creation of a variety of conceptual frameworks at the international level (Aday et al., 2004; Arah et al.,

2006; Atun & Mendabde, 2008; Commonwealth Fund, 2006; Hurst & Jee-Hughes, 2001; IHP,2008; Jee & Or, 1999; Kelley & Hurst, 2006; Klassen et al., 2009; Murray & Frenk, 2000; Roberts et al., 2008; Sicotte et al., 1998). While these frameworks have varied purposes they all aim to provide a better understanding of what a health system is, its goals, and the underlying structure and factors that drive its performance. Indeed careful examination of the available international frameworks suggests that over time there is a degree of convergence, both in the framework architecture and goals but also in the problem areas they encounter. This suggests that the gains from a creating a new framework have progressively decreased in proportion to the efforts required to undertake such a task. The priority now is to clarify areas where there are long-standing differences in matters of understanding, focus and principle.

One of the main areas of debate that needs to be addressed is determining where the boundaries of the health system lie. There can be no right answer to this question, as there are sound reasons for promoting the use of both wider and narrower boundaries. Narrow boundaries are better suited to holding stakeholders accountable, while broader boundaries are better for a more holistic understanding of determinants (Figure 1). Moreover, boundaries that correspond to the goals of the entity undertaking HSPA will be more successful at engaging key stakeholders in the HSPA process and thus more likely to ensure sustainability of these efforts in the long run.

Figure 1: Performance measurement implications of setting health system boundaries



Yet, lack of consensus on this issue makes comparisons across frameworks and national performance assessments difficult. A possible solution to this problem may be to explicitly identify different levels of the health system and attempt to measure the contribution of each level towards the achievement of the performance goals. This is done in the OECD Health Care Quality Indicators framework (Arah et al., 2006) where the representation of the health system through four tiers allows health care to be placed within a broader conceptualization of the health system as well as the economic, social and political context of countries. While acknowledging that health care is only one of the wider determinants of health, this structure allows health care performance to be measured, without being subsumed within the wider health system model. Indeed, it is advisable that regardless of how boundaries are set, an HSPA programme should measure factors that influence the key health system goals in order to understand how much of the observed performance is amenable to health system influence, while also providing stakeholders with enough information to advocate for intersectoral action. This solution also allows users to adapt universal measures and comparisons to the specific design and priorities of their own health system. We see, for instance, a growing focus on well-being (the measurement of which is now in development in various agencies, including WHO), and this may or may not feature in a given performance assessment depending on the priority accorded it and the boundaries set for the health system.

There is a growing consensus as to the main components of a health system. However, the definition will necessarily vary between countries depending on the institutional arrangements in place. We nevertheless identify five key elements common to all systems, namely: service provision, financing, resource generation, leadership/governance and risk factors, which we believe should be included in any framework in order for it to provide a rounded picture of the organizational structure within which health systems operates. A clear understanding of how these elements relate to performance is necessary in order for users of the HSPA to relate performance back to their own system and identify what policy levers can be addressed to secure improvement.

Unlike the issue of boundaries, there seems to be relative consensus on the goals of a health system. However, there are still differences in interpretation as to what these goals encompass. In particular, concepts such as Responsiveness, Quality, Equity and Efficiency tend to have a variety of connotations. There is also a lack of consensus as to how the different goals are related to one another. For example, is access to health services an aspect of equity or responsiveness? Is efficiency included in what we mean by quality or is quality part of efficiency? This ambiguity leads to a lack of clarity that makes the operationalization of these frameworks difficult and controversial (Papanicolas and Smith, forthcoming).

In order for the assessments of different domains of performance to become comparable across systems and organizations, there is an urgent need for consensus on definitional issues.

The key unresolved question for health system comparison in relation to system goals is whether to treat “efficiency” as the overarching goal of the health system, within which all comparison is to be embedded, or whether to adopt a more limited goal of offering fragmentary indicators of productivity, for example in the form of unit costs of individual services. The advantages of the former are that it offers a coherent intellectual framework, and that many of the inputs to the health system (such as manpower) are easiest to measure at the whole-system level. The disadvantage is that whole-system measures offer little diagnostic information on where inefficiencies are arising.

3 How to compare key domains of performance

Performance measurement evaluates the extent to which a health system meets its key objectives. Most HSPA efforts focus on common dimensions of measurement, such as health improvement or health status, responsiveness, equity and efficiency. Population health examines both aggregate data on the health of the entire population and health service outcomes, which focus on the outcomes for patients. Responsiveness considers dimensions unrelated to health outcomes such as dignity, communications, autonomy, prompt services, access to social support during care, quality of basic services and choice of provider. In each of these domains both the average attainment of health systems in each domain (effectiveness) and their distribution across the population (equity) are of interest. Often HSPA efforts also identify financial protection as an explicit health system goal. Finally, although its exact characterization is subject to debate, some concept of efficiency is usually an important dimension of performance. Table 1 considers some of these key dimensions of HSPA efforts, considering why they are important to measure and what the key areas of comparison are for policy-makers.

Progress in the development of data collection techniques in the different dimensions of health performance measures has been variable. Some areas, such as population health, can be quite reliably captured through established indicators, while other areas, such as efficiency, are in earlier stages of development. Moreover, some dimensions of health systems are intrinsically hard to capture due to their abstract or contested nature, such as multifaceted concepts like responsiveness. Clarification of the concepts to be assessed is therefore the first task of comparison. In terms of interpreting and constructing measures it is then essential to be aware of the main measurement instruments being used, whether there are gaps or disputes in the existing approaches,

Table 1. Performance measurement implications of setting health system boundaries

Dimension	Motivation for international comparison	Areas of interest for comparison
Population health	<ul style="list-style-type: none"> • To facilitate a comparison of health within and across countries considered from a broad aggregated perspective, which includes their contributions to many of the risk factors for disease as well as to the delivery of health care. • To facilitate a comparative assessment of how health systems contribute to population's health. 	<ul style="list-style-type: none"> • Life expectancy • Mortality by age group and condition • Morbidity • Avoidable mortality • Population risk factors (including some measures of inequalities to enable comparisons) as predictors of future population health
Health service outcomes	<ul style="list-style-type: none"> • To facilitate a comparative assessment of how health services assist individuals in realizing their potential health. 	<ul style="list-style-type: none"> • Performance of different areas of the health system (preventative care, primary care, secondary care, long-term care, mental health) • Health system outcomes • Health system processes
Equity	<ul style="list-style-type: none"> • Allows an assessment of inequalities in health among different population/ demographic/social groups within and between countries. • Allows an assessment of inequalities in access and/or utilization of services among different population/ demographic/social groups within and between countries. • Allows an assessment of inequalities in financing of health services among different population/ demographic/social groups within and between countries. • Allows an assessment of inequalities in responsiveness of health services among different population/ demographic/social groups within and between countries. 	<ul style="list-style-type: none"> • Distribution of health status by population/ demographic/social groups • Distribution of access/ utilization of health services by population/ demographic/social groups • Progressivity of financing system • Distribution responsiveness of health services by population/ demographic/ social groups

Table 1. Performance measurement implications of setting health system boundaries (continued)

Financial protection	<ul style="list-style-type: none"> • To enable a comparative assessment of how the health system protects citizens from the financial consequences of ill health. 	<ul style="list-style-type: none"> • Out-of-pocket spending • Catastrophic expenditures on health care • Impoverishing expenditures on health care • Fairness of financing
Responsiveness	<ul style="list-style-type: none"> • To facilitate a comparative assessment of how satisfied health systems leave the patients with whom they come into contact. 	<ul style="list-style-type: none"> • Patient satisfaction • Patient choice • Respect of patients dignity • Prompt attention to medical needs
Efficiency	<ul style="list-style-type: none"> • To facilitate a comparative assessment that allows policy-makers to pinpoint which parts of the health system are not performing as well as they should be, based on the experiences of other health systems. 	<ul style="list-style-type: none"> • Value for money of services • Waste of resources • Effective coverage • Disease costs

and how useful current indicators are for the purpose of assessing system performance. Ultimately, the collection of information is useful not only for comparative purposes but also to identify key areas of weakness and strength internally. This section considers some of the key measures for each performance domain as well as the related policy uses and abuses – that is, what the indicators can and cannot tell us about system performance.

a. Population health

Without question the main aim of any health system is to improve the health of the population that it serves. Thus population health is often the first area considered when evaluating the performance of a health system, requiring aggregated data on the health status and health improvement of the population. Principal indicators in this area include measures such as life expectancy at particular ages, age-standardized mortality, premature or infant mortality, years of life lost, disability-adjusted life-years (DALYs) – all of which capture generic information on population health. These types of measures take a broad perspective, which measures the effect on the health of the population of many risk factors for disease as well as the delivery of health care. This

perspective can be attractive from a political point of view because it demonstrates the role that broader determinants of health play in determining health status. However, it also creates major methodological challenges in seeking to attribute changes in health to any particular policy.

Consequently, without discarding the broader approaches, more recent research has focused on measuring the contribution of health care to improved health. This has led to the development of concepts such as avoidable mortality and the use of tracer conditions. Avoidable mortality, in its broadest sense, includes deaths considered to be avoidable by use of appropriate and timely medical care and those preventable by population-based interventions. It can be further broken down into subsets of amenable mortality which refer to conditions where “it is reasonable to expect death to be averted even after the condition develops” and preventable mortality which includes deaths from conditions that can be prevented by population-based interventions, but where the contribution of health care may be limited once the condition has developed (Nolte & McKee, 2004). Recent work in this area considers how avoidable mortality can be measured in different countries, and how metrics can be used in comparative analyses (Box 2).

Box 2. Avoidable Mortality in European Health Systems (AMIEHS)

Funded under the European Union Public Health Programme, the AMIEHS project, led by Erasmus Medical University and coordinated jointly with the London School of Hygiene & Tropical Medicine, brings together partners in seven EU countries with the aim of creating better comparable indicators with which to measure the contribution health care makes to population and how this varies among countries. The project aims to develop a set of avoidable-mortality-based indicators that can be used in future surveillance of the performance of health systems in Europe.

The project aims to undertake the following initiatives which will assist policy-makers and researchers with the understanding, measurement and use of avoidable mortality indicators in Europe:

- to conduct a systematic review of the literature to assess to what extent causes of death can be considered avoidable;
- to gather in-depth information on the introduction of medical innovations in seven countries;
- to develop a set of avoidable mortality-based indicators that is agreed upon; and
- to prepare an electronic atlas of avoidable mortality in 25–30 countries in Europe. (http://survey.erasmusmc.nl/amihs/maps/J45_J46/atlas.html)

Source: <http://amihs.lshmt.ac.uk/>

Tracer conditions are based on the premise that carefully selected health problems can provide insights into the more general performance of elements within the overall health system (Nolte, Bain & McKee, 2009). Table 2 considers the main types of indicators used to measure population health as well as their policy uses and abuses; what the indicators can and cannot tell us about system performance.

As evident from Table 2, many of the existing measures of population health fail in their simplest form to distinguish the contribution of health care from extraneous factors. Moreover, all indicators still suffer from a number of methodological problems that need to be addressed in order to make international comparisons more meaningful. Important issues involve availability and coding of data, particularly of data on cause of death where there are

Table 2. Main indicators for population health

Main indicators	Policy uses	Limitations
<i>Generic indicators:</i> e.g. life expectancy, age-standardized death rates	These are broad indicators of health service delivery and achievement of desired population health outcomes.	<ul style="list-style-type: none"> • Broad indicator of health; • Mask contributions of specific causes; • Exclude morbidity; • Need further disaggregation by age and cause.
<i>Age/disease-specific indicators:</i> e.g. infant mortality rate, perinatal mortality; age-specific mortality, cancer five-year survival	Allow more detailed analysis of specific outcomes of the quality of health care.	<ul style="list-style-type: none"> • Susceptible to variations in recording and reporting practices. • Rely on precise definitions not always adhered to in practice (perinatal mortality). • Capture influence of broader health determinants. • Are based on small numbers. • Complex interpretation of underlying causes. • Need to be interpreted in context of risk factor and disease prevalence, and policies in other sectors. • (Cancer survival) has to be viewed alongside mortality and incidence rates.

Table 2. Main indicators for population health (continued)

Main indicators	Policy uses	Limitations
<i>Morbidity indicators:</i> e.g. health survey data, incidence notifications, health service utilization	At present morbidity data is of limited use in assessing the contribution of health care to population health but there is great interest in emerging initiatives to develop registries for specific conditions, which focus on health outcomes and service delivery aspects.	<ul style="list-style-type: none"> • Reporting bias (health survey data). • Non-specific to health care interventions (health survey data). • Variations in notification requirements and diagnostic practices. • Data coverage (often excludes private sector). • Representativeness (utilization only shows people who accessed the health service).
<i>Summary indicators:</i> e.g. Health-adjusted life expectancy (HALE), DALYs	Allow a more rounded picture of population health assessment that takes into account both mortality and morbidity of populations.	<ul style="list-style-type: none"> • Controversial methodology (age and disability weightings). • Limited availability of the required health status data, especially over time.
<i>Indicators measuring the contribution of health care:</i> e.g. avoidable mortality (AM), tracer conditions	Selected mortality-based indicators combined with supplementary techniques (such as tracer methodology) allows the exploration of the individual aspects of health service delivery process and highlight potential gaps and weaknesses.	<ul style="list-style-type: none"> • Aggregate measure requiring further disaggregation (AM). • Variations in list of amenable causes and age limits (AM). • Time lags for outcomes of specific interventions (AM).

Source: Adapted from Karanikolos et al., forthcoming.

problems of comparability among countries and over time. Coding is also influenced by changes among versions of the International Classification of Diseases and national coding rules, often reflecting whether automated coding is used or not. There are also large gaps in the availability of evidence on the effectiveness of treatments in reducing mortality, which is rarely an outcome of randomized trials. Especially in Europe, there are no internationally comparable surveys of clinical practice and health care experience. In the future, more investments should be made in the development of internationally comparable clinical databases providing risk-adjusted information on individual outcomes of treatment. In the meantime, policy-makers looking to understand changes in health status of their population might consider making greater use of tracer conditions, through which the everyday experiences of those in need of care can be understood and addressed.

Not all of the problems facing population health indicators can be readily resolved, and so it is important to understand limitations and take them into account when interpreting data. Some of the most important issues to consider in the interpretation of data are the difficulty in attribution, not only from other factors that can influence health status, but also from distinguishing changes in the ability of health systems to prevent death once disease has occurred from changing incidence of that disease, thus complicating attribution. Moreover, observed changes in mortality from particular causes, even if not due to artifact, can reflect changes in any one, or a combination of innovation, coverage or quality. These can be difficult to distinguish. Finally, in the analysis of national policy, it is important to consider the varying and often diffuse time lags between the introduction of a policy or treatment innovation and a change in outcome.

b. Health services outcomes

Although the main objective of health services is ultimately to assist individuals in realizing their potential health and thus promote the health of the population, measuring the contribution of health services to health outcomes involves quite distinct challenges. For example it is essential to ensure that the services being compared are directly comparable, and that proper adjustment is made for differences in the populations being served. Direct indicators of the contribution of health services to health status are available in the form of health service quality measures, such as standardized hospital mortality rates and numerous disease-specific health outcome measures. To date, outcome measures in widespread use capture only a limited set of dimensions of the broad “health concept” and tend to focus on mortality (case fatality rates, hospital standardized mortality rates) adverse events and complications (patient safety indicators) as well as readmissions and avoidable admissions (hospital

care, mental health care, primary care). Outcome measures in areas such as disabilities and discomfort are far less prevalent. For the purposes of international comparison in high-income settings, the OECD quality indicators project is an important resource (Box 3). While such measures offer some indicators of the performance of individual organizations (after suitable adjustment for case-mix and other contextual circumstances), international comparison can be complicated by different organizational settings and reporting conventions.

Box 3. Health Care Quality Indicators (HCQI)

The HCQI project, led by the OECD, was initiated in 2001 with the aim of measuring and comparing the quality of health service provision in the different OECD countries. Over the years the HCQI project has grown into a robust source of internationally comparable data on the quality of care as well as a forum for policy-makers and researchers to assist in the improvement of quality measurement. The project has focused on producing comparable indicators in key areas of health care such as primary care, acute care, mental health care, cancer care, patient safety and patient experience.

Most data are collected from administrative databases, registries and population surveys. Following the compilation of the data considerable efforts are undertaken to methodologically refine the data assessment and collection procedures, such as assessment of data quality, refinement of technical specifications, enhanced data collection guidelines and questionnaires and the harmonization of approaches to age/gender standardization. Currently there are nearly 40 indicators that are routinely collected and reported for Member States every two years.

Source: OECD, 2010.

Routine use of patient-reported outcome measures (PROMs) has recently been introduced in England, and experience there will indicate the potential they offer in addressing some of the existing measurement gaps. The uptake of PROMs for routine purposes has hitherto been hindered by comparability issues especially in content validity and the relative importance of different criteria. Some disease-specific questionnaires upon which PROMs are based are not only considered costly and time-consuming but may also be intrusive and burdensome to patients, potentially jeopardizing the professional–client relationship. However, simpler generic instruments such as the EQ5D in England have secured high response rates both before and after surgery.

In many performance measurement initiatives, measures of health care process are used in preference to more direct measures of outcome. These have the virtue of administrative convenience, can be measured immediately, and are

easier to attribute directly to the efforts of the health services. Furthermore, they reflect compliance with what is considered good practice and may therefore be a better measure of assessing the quality of health care providers when the processes measured are known from research evidence to lead to good patient outcomes. In contrast, more distant outcome measures may be influenced by many factors other than the quality of care alone. However, process measures may ignore the ultimate effectiveness or appropriateness of the intervention, and their use pre-judges the nature of response to a health problem, which may not be identical in all settings. Judging health services by adherence to standards or guidelines based on single diseases can therefore be misleading, especially in situations where health services are dealing with patients with multi-morbidities.

One of the main methodological issues in the development of outcome indicators is data quality and availability. Any further development in measuring health outcomes therefore depends on the enhancement and expansion of current data sources. This is likely to entail more widespread use of Unique Patient Identifiers to link various data sources, the use of secondary diagnoses codes, present-at-admission codes and standardization of procedure codes. This is a prerequisite for increasing the potential for case-mix adjustments for outcome measures and the further development and testing of new measures such as PROMs and other measures that can use Electronic Health Records and patient surveys as their main data source. Contingent on privacy and data protection regulation and supported by the necessary research and development to test reliability and validity, health outcome measures, capturing health services outcomes, will become an increasingly important part of health system performance assessment.

Table 3. Main indicators for health service outcomes

Main indicators	Policy uses	Limitations
<i>Hospital outcome indicators:</i> e.g. Hospital Standardized Mortality Rates (HSMRs), case fatality rates for Acute Myocardial Infarction (AMI) and stroke, patient safety indicators and hospital readmission rates	Hospitals are considered by many policy-makers to be the epicentre of the health care system – these indicators consider the contribution hospitals make to health outcomes over time.	<ul style="list-style-type: none"> • HSMRs do not account for preventable deaths and the observation that a majority of deaths are unavoidable. • Differences across hospital systems and records make comparability across hospitals and countries difficult. • Lack of data collection at the individual level, and lack of supporting information for which to case-adjust indicators pose issues for reliability and comparability of indicators. • Readmission indicators pose problems of comparability due to different definitions of time frames and type of readmission investigated as well limitations in case-mix. • More information is needed to determine the actual relationship between readmissions and the quality of care.
<i>Patient-reported outcome indicators (PROMs):</i> e.g. SF-36, EQ5D	PROMs are useful as they are not only able to capture and regularly assess aspects of health that are of most concern to patients but are argued to be essential for the assessment of patient need and communication between patient and provider in routine care.	<ul style="list-style-type: none"> • Individualized instruments are very time-consuming and often involve complex interviews. • Uptake largely hindered by comparability issues, especially in content validity and the relative importance of different criteria. • Questionnaires are considered costly and time-consuming. • May be regarded as “soft information” by some stakeholders.

Table 3. Main indicators for health service outcomes (continued)

<p><i>Indicators for long term care:</i> e.g. process indicators, client experiences, nursing-related outcomes (bed-sores, patient falls), outcomes of targeted illnesses (diabetes, dementia, etc.)</p>	<p>Measuring the quality of long-term care services is of high importance given the global ageing population trend and the present morbidity and disability patterns associated with chronic diseases.</p>	<ul style="list-style-type: none"> • Complexity of chronic care, and different levels and settings of service provision hamper the identification of possible quality indicators and create difficulties in standardized data collection and reporting conventions. • It is hard to pinpoint medical outcome measures to the performance of long-term care institutions, as a result nursing-related indicators are usually the only ones used. • Outcomes of targeted illnesses (such as avoidable admissions for diabetic patients) may be as much an indicator for other areas, such as GP or specialist care.
<p><i>Indicators for primary care:</i> e.g. avoidable events, preventable admissions, process indicators</p>	<p>The significance of primary care lies in its effectiveness in preventing illness and death and in its association with a more equitable distribution of health in populations.</p>	<ul style="list-style-type: none"> • The wide variation in payment and contractual organization for primary care services across countries inevitably translates into differences in the scope of data collection possible. • Despite progress in the international collection of data, the most robust source for deriving indicators in primary care remains to be hospital administrative data. As such, they do not provide a complete assessment of a primary care system's quality of care. • Collection of avoidable admissions and adherence to processes relating to specific clinical areas (e.g. diabetes, asthma, COPD) are sufficiently relevant to policy and scientifically sound for potential use in international data collection but still of limited availability across countries.

Table 3. Main indicators for health service outcomes (continued)

Main indicators	Policy uses	Limitations
<i>Indicators for mental health:</i> e.g. rates of unplanned readmission for schizophrenia and bipolar disorder	Mental health problems are common, affecting all sections of society and every age group.	<ul style="list-style-type: none"> • Variation of organization across countries makes the assessment of the quality of mental health care services for evidence-based policy difficult. • The availability of national indicator data suitable for international comparison is extremely limited due to the complex nature of mental health disorders, the differences in diagnostic and therapeutic practices, institutional government barriers as well as differences in the coding and reporting of mental health care within and between countries.
<i>Indicators for preventative care:</i> e.g. screening rates	Screening is of great interest to policy-makers because of the significant bearing it may have on survival prospects.	<ul style="list-style-type: none"> • There are many methodological issues relevant to the data collection for and comparability of cancer screening indicators in combination with other cancer outcome information such as five-year survival rates and cancer mortality rates, including data sourcing (e.g. surveys versus registries), heterogeneity in cancer survival and screening reporting periods, age standardization, the extent to which country data is nationally representative, and, perhaps most importantly, a lack of cancer staging data.

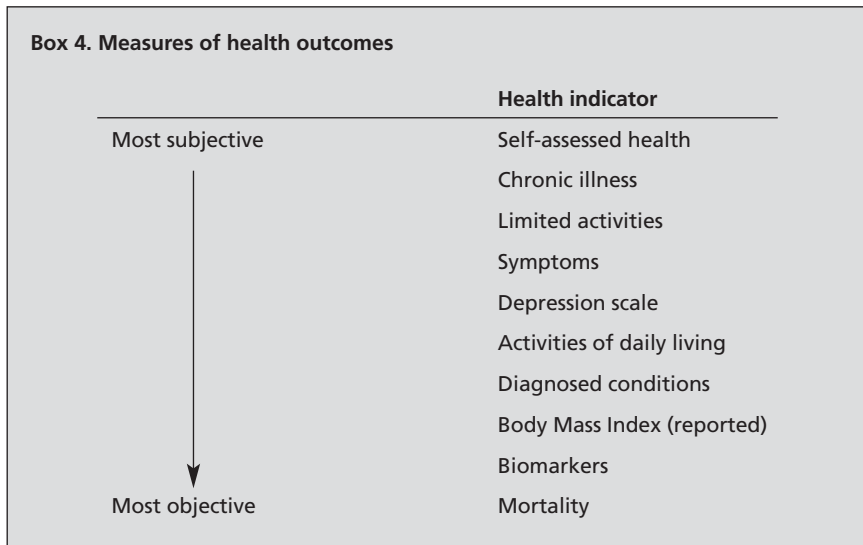
Source: Adapted from Klazinga & Li, forthcoming.

c. Equity

The exact specification and importance of equity objectives is for individual nations to determine. We are concerned with ensuring that stakeholders are aware of the tools to use in order to measure their performance with regards to the policies they have chosen to pursue.

The principle of equity in health is the principle of equal (or equitable) health outcomes (for example, quality-adjusted life expectancy). When analysing inequalities in health outcomes at the individual level, different health outcome measures have been used in the literature, from subjective measures of health, such as self-assessed health, to more objective measures of health such as biological markers (Box 4). Summary measures of population health can be used, such as: disability-free life expectancy (DFLE), HALE and DALYs by socioeconomic status (SES). However, data linkages often make this approach problematic. Second, it is possible to apply mortality-based indicators, such as life expectancy, total mortality rate and infant mortality rate by SES, which requires linkage between death registries and socioeconomic status information. Third, morbidity-based indicators such as self-rated health by SES or self-rated disability by SES can be used, feasible for international comparison studies as relevant data are often included in health surveys.

Various methodologies can then be used to investigate disparities in health among different groups of the population. When investigating socioeconomic



inequalities in health, there are several groups of methods that can be applied to determine absolute inequalities in health, such as the Gini coefficient. Alternatively, regression methods can be used to derive measures of association, such as the odds ratios. Finally, more advanced methods that create indices based on ranking of socioeconomic variables may be used, such as the Concentration Index. This decomposes overall inequalities in health by the main contributory need and non-need variables (Wagstaff, van Doorslaer & Watanabe, 2003; O'Donnell et al., 2008). This methodology can be extended further to take into account the longitudinal perspective, by measuring long-term income-related inequalities in health (Jones & López-Nicolás, 2004).

Yet research on health equity is concerned not only with equal or equitable health outcomes, but also with equity in health care utilization/access, health care financing and responsiveness (O'Donnell et al., 2008). In the case of equity in access or utilization to health care, the main interest focuses on measuring whether individuals with the same level of clinical need are receiving the same level of health care, a horizontal equity principle. Many of the measurement challenges in this area derive from difficulties in conceptualization and measurement of the concepts of access and need (Allin, Hernández-Quevedo & Masseria, 2009). The main methods used for measuring inequity in access to health care for cross-country comparison studies are: a simple comparison of rates of access for different groups of the population, the use of regression methods such as adjusted odds ratios, which are a measure of association, and the Gini-like coefficients such as the horizontal inequity index.

Financing systems can take three broad forms: progressive (payments are an increasing proportion of ability to pay, such as income taxes), proportional (payments represent a constant proportion of ability to pay, such as payroll taxes) and regressive (payments are a decreasing proportion of ability to pay, such as out-of-pocket payments). When measuring equity in finance, the interest focuses on the extent to which health care payments are related to ability to pay. In particular, the vertical dimension of equity is fundamental, in order to understand the extent to which those with unequal ability to pay do pay differently for health care. In order to measure vertical equity, progressivity indices (Wagstaff et al., 1999) have been designed, such as Lorenz curves and Gini coefficients, Concentration curves and the Kakwani index.

Responsiveness of the health care system has been defined as the extent to which health services are aligned with user preferences in domains such as patient autonomy, choice and quality of amenities (see section e, Responsiveness). Methods are being developed to measure systematic differences in responsiveness by social group (see Valentine et al., 2009). The main challenge is the continuing debate over how to define and conceptualize responsiveness, as discussed below.

Table 4. Main indicators for equity

Main indicators	Policy uses	Limitations
<p><i>Equity in health outcomes:</i> i.e. distribution of health among different groups. Constructed from indicators of health status and socioeconomic variables.</p>	<p>The main goal of the health systems would be to decrease the gap between health outcomes that are systematically related to socioeconomic status and hence, focusing on reducing socioeconomic inequalities in health outcomes is the usual performance measure in the health systems. There is also substantial evidence on inequalities in health across many population groups (region, ethnicity, gender, language).</p>	<ul style="list-style-type: none"> • Measures are limited by the availability of outcome indicators available as well as their linkage to socioeconomic variables. • Numerous surveys include outcome and socioeconomic information, but the reliability on subjective measures of health status raises important methodological challenges that relate to the potential reporting bias that could appear. • Longitudinal data is lacking, especially for a comparative analysis, making it difficult to understand changes over time. • Objective measures such as physicians' assessments or hospital stays are best for comparative purposes. However, the availability of objective measures of health, such as biomarkers, is limited. • Biomarkers may still be subject to bias and are not included in longitudinal data. • Standardization of biomarker data collection across countries is also an issue.
<p><i>Equity in access or utilization:</i> i.e. distribution of access or utilization of health care for equal amounts of need among different groups. Constructed from indicators of access, utilization, need, unmet need and socioeconomic variables.</p>	<p>The main focus of equity of access or utilization is to measure whether individuals with the same level of need are receiving the same level of health care, and thus to determine if inequalities are unavoidable.</p>	<ul style="list-style-type: none"> • Often the terms "access" and "utilization" are used interchangeably, implying that an individual's use of health services is proof that he/she can access these services. However, utilization is not equivalent to access. • Utilization and need are often captured by survey information, which can suffer from reporting bias as well as comparability issues across countries. • Little data is collected longitudinally, and there are large gaps on data that inform on environmental factors.

Table 4. Main indicators for equity (continued)

Main indicators	Policy uses	Limitations
<p><i>Equity in financing:</i> i.e. distribution of financing of health care among different groups. Constructed from indicators of expenditure and income.</p>	<p>The main focus on equity of financing is to determine whether the health system is progressive (namely, how much larger payments are as a share of income for the poor than for the better-off).</p>	<ul style="list-style-type: none"> • The tabulation of average incomes and health care payments by income groups have been used in the literature. However, this methodology is not able to establish how much more progressive is one system than other.
<p><i>Equity in responsiveness:</i> i.e. distribution of responsiveness among different groups. Constructed from information on responsiveness and socioeconomic indicators.</p>	<p>Responsiveness of the health care system has been defined as the extent to which health services are aligned with user preferences in domains such as patient autonomy, choice and quality of amenities. Equity in this area considers the distribution of this health system goal across groups in the population.</p>	<ul style="list-style-type: none"> • There is still lack of clarity on what is meant by responsiveness, resulting in numerous different, conflicting measures within and across countries.

Source: Adapted from Hernández-Quevedo, Papanicolas & Machenback, forthcoming.

d. Financial protection

Financial protection is a multidimensional concept that usually refers to the extent to which people are protected from the financial consequences of ill health (WHO, 2000). Such protection is a key objective of health systems worldwide. The most appropriate set of policies for improving financial protection will depend on the particular context. It is thus of foremost importance to have accurate means of measuring the extent of financial protection and of understanding the determinants of financial risk in a particular health system, so as to aid the identification of suitable policy levers. Useful measurement tools have been developed with this aim.

Financial protection can be observed in several forms, such as high out-of-pocket payments for health care or lack of access due to financial barriers. The multidimensional nature of financial protection has made it difficult to develop a single indicator capturing the full extent to which people are financially protected from health shocks. Instead, partial measures of the incidence and magnitude of households' direct payments for health care usually form the basis of metrics for financial protection assessment and system comparisons. The emphasis so far has been on the incidence of (1) "catastrophic" health payments, relative to some threshold of household income, and (2) "impoverishing" health payments, relative to some pre-defined poverty line. More recent literature has focused on the incidence and magnitude of user payments for health care, mainly out-of-pocket expenses (and sometimes only these), to get at the issue of the extent of financial protection in a health system.

Such metrics, based on information provided by household surveys, fails to recognize that inability to pay may deter access to necessary care, hence resulting in very low or zero health expenditures reported (Moreno-Serra, Millett & Smith, 2011). The typically right-skewed distribution of health care spending observed in household surveys (with a high number of zeroes) underlines, among others, the problem of inadequate risk-pooling, and suggests that many households worldwide experience major financial barriers securing access to health services. While this may be linked to equity of access, it is also an important indicator of lack of financial protection *per se*. In addition, conventional measures of financial protection for performance assessment cannot say much about the specific drivers of financial risk in a health system. This has led to suggestions that some indicators of coverage for certain basic health care treatments should be used to complement the information provided by currently used financial protection metrics (see, for instance, WHO, 2010).

Cross-country studies have normally focused only on comparing the measured incidence of catastrophic or impoverishing spending. The limited scope of such studies has meant that conclusions about system-wide determinants of differences in financial protection levels across countries have often been based mostly on descriptive or anecdotal evidence.

Table 5. Main indicators for fairness in financing

Main indicators	Policy uses	Limitations
<i>Catastrophic and impoverishing health payments</i>	<p>Catastrophic spending indicators can offer a useful picture of the extent to which citizens in a health system suffer hardship due to the costs of health care services.</p> <p>Impoverishing health payment indicators allow system comparisons in terms of the number of people being pushed into poverty, relative to a minimum living standard, due to illness.</p>	<ul style="list-style-type: none"> • Provide only limited insights about what the major determinants of inadequate financial protection are in a given context. • Do not inform as to whether factors related to financial barriers of access to health care play a relevant role for the measured extent of financial protection and, if so, which individuals are more affected by such barriers. • Lack of work investigating the various aspects related to access to health services as determinants of financial protection levels means that the comparison of incidence of catastrophic or impoverishing spending across countries can only result in speculative conclusions about system-wide determinants of differences in financial protection levels across countries.
<i>Out-of-pocket payments</i>	<p>A simple strategy to gain some insight into how far citizens in a health system are protected against the financial consequences of illness is to look at the contribution of private health spending to the financing of the system.</p>	<ul style="list-style-type: none"> • Cross-country examinations of the relative importance of out-of-pocket expenses for funding the health system can convey helpful insights for performance comparisons of financial risk. • Measuring and comparing the actual extent of financial protection across health systems would require the analyst to examine micro-data relating households' out-of-pocket health expenses to some metric in terms of their living standards.

Table 5. Main indicators for fairness in financing (continued)

<i>Index of fairness of financial contribution (WHO, 2000)</i>	The notion of fairness of financial contribution developed in the <i>World health report 2000</i> is based on the premise that a fair health system ensures that households make health care payments according to their ability to pay rather than risk of illness, hence being protected against the risk of falling into poverty – or being deterred from seeking care – due to health care costs.	<ul style="list-style-type: none"> • The indicator is unable to discriminate between countries where health payments are progressive or regressive; the extent to which inequalities are due to horizontal inequity, vertical inequity; and between different proportions of national income going to the health care system.
--	---	--

Source: Adapted from Moreno-Serra, Thompson & Xu, forthcoming.

e. Responsiveness

The *World health report 2000* on the performance of health systems proposed responsiveness to citizens' expectations as a central and distinct goal. It pushed forward a debate that frames responsiveness as a valued and desired outcome of health system interventions regardless of the extent to which those interventions lead to health improvement (WHO, 2000). Health services reforms in many countries have been placing increased explicit emphasis on improving responsiveness to patients and increasing both population and patient satisfaction.

In preparation for the *World health report 2000*, an extensive literature review covered disciplines including sociology, anthropology, ethics, health economics and management in order to elicit what people value most in their interactions with the health system (De Silva, 2000). This was used to select a common set of seven dimensions (or domains) that characterize the concept of responsiveness, which was subsequently augmented by an eighth, as summarized in Table 6. Four are grouped under "respect for persons" (dignity, confidentiality, autonomy and clarity of communication) and four under "client orientations" (prompt attention, quality of basic amenities, access to family and community support, and choice of health care provider).

Table 6. WHO dimensions of responsiveness and questions used to measure it in the World health report 2000 and two WHO population surveys

Dimension	<i>World health report 2000: grouping and weighing (results based on interviews with experts)</i>	Multi-country survey study 2000/2001	World health survey 2002
<i>Respect for persons</i>			
Dignity: Respectful treatment and communication	16.7%	4 questions	2 questions
Confidentiality of personal information	16.7%	2 questions	2 questions
Autonomy: Involvement in decisions	16.7%	3 questions	2 questions
Clarity of Communication	Not included	4 questions	2 questions
<i>Client-orientation</i>			
Prompt attention: Convenient travel and short waiting times	20%	2 questions	2 questions
Quality of basic amenities: Surroundings	15%	3 questions	2 questions
Access to family and community support: Contact with outside world and maintenance of regular activities	10%	3 questions	2 questions
Choice of health care provider	5%	3 questions	1 question

Source: Busse, forthcoming.

There is still uncertainty as to what dimensions are included in the domain of responsiveness, and this lack of clarity leads to variation in the areas measured. The consequent variations in results reflect the different approaches towards the selection of weights, domains and indicators (i.e. surveys capture different phenomena); differences in the methodology of data collection (e.g. sampling)

and interpretation; as well as actual differences in health systems' levels of responsiveness.

Both responsiveness and satisfaction are terms that aim to capture the degree to which health systems, or their components, are successful in responding to the expectations of the general population or a population subgroup of patients. A wide range of methods has been used to attempt to measure responsiveness and/or satisfaction over the last decades, most visibly work by Blendon et al. (1990); population satisfaction questions in Eurobarometer surveys since 1996 (European Commission, 1996, 1998, 1999, 2000, 2002); the Picker Institute's development of patient experience surveys (Coulter & Cleary, 2001, Jenkinson, Coulter & Bruster, 2002); the EUROPEP instrument to assess general practice (Grol et al., 2000); the *World Health Report 2000* (WHO, 2000), as well as work by the Commonwealth Fund (Schoen et al., 2007).

Responsiveness and satisfaction are distinct but related concepts. The WHO definition refers to "responsiveness to the legitimate expectations of the population for their interaction with the health system". This implies that there can be illegitimate or unjustified expectations too, but the instrument only captures those that are regarded as legitimate. The "satisfaction of the overall population with the health system" may be influenced by other expectations (which experts or policy-makers may consider illegitimate) and factors outside the direct control of the health care system. Satisfaction is likely to be more dependent on expectations than responsiveness surveys, so higher satisfaction levels may to some extent be a function of lower expectations. WHO used a vignette approach in its *World Health Survey* in order to correct for different expectations, but this approach was dropped due to the complex data and analytic requirements (Rice, Robone & Smith, 2010).

In principle, the concept of population satisfaction with the whole health system is straightforward. In fact, it is difficult to measure satisfaction as the answers to all questionnaires depend on the specific wording of the question asked as well as the answer categories provided. The answers depend particularly on factors not yet well-understood, that is (1) the context in which a survey takes place, for example, coloured by recent media coverage of scandals, fraud or underprovision of services; (2) no differentiation between the system as a whole and certain subsectors about which the respondent may be more knowledgeable; or (3) the inability to differentiate between the health care system and government in general. Busse et al. (2011) provide an overview of different population surveys used over recent decades.

Another source of information in this area is surveys of patients' experience with treatment by particular providers. Such surveys relate more to responsiveness than to satisfaction as they are based on (1) predetermined

domains and (2) patients' actual health service encounters. Widely used surveys are available for inpatient, outpatient, general practitioner, maternity, mental health, community and emergency care, as well as for specific groups of patients, for example those with diabetes, heart disease or cancer.

In summary, different questionnaires with different items in relation to responsiveness sometimes lead to inconsistent or contradictory results and may be difficult to interpret. Overall, the lack of consensus on concepts and metrics makes it hard to draw clear conclusions on the differences between health systems, and even less about the health system strategies that may explain them. Progress in this area needs to be made initially on the conceptual front and, once more clarity is obtained, regarding what needs to be measured before more consistent indicators can be produced within and across countries.

f. Efficiency

Efficiency indicators serve as a summary measure of the extent to which the inputs to the health system, in the form of expenditures and other resources, are used wisely to secure the goals of the health system. Economists often make a distinction between two types of efficiency: allocative efficiency and technical efficiency. Allocative efficiency indicates the extent to which limited resources are directed towards producing the correct mix of health care outputs in line with the preferences of payers.

Indicators of allocative efficiency exist at the micro-level (are the "right" treatments being used) or at a macro-level (is the appropriate level of resources dedicated to health care, relative to other sectors of the economy, given prevailing societal values). Allocative efficiency can also be considered within the health care system at an intermediate level to examine whether the correct mix of services is funded, such that attainment of system objectives such as health outcomes is maximized for a given aggregate level of expenditures.

In contrast, technical efficiency indicates the extent to which the system is minimizing costs in producing its chosen outputs, regardless of the value placed on those outputs. Whereas allocative efficiency assesses whether the system is producing an appropriate mix of outputs, technical efficiency makes no judgement on how much the outputs are valued by society, and is concerned solely with any "waste" in creating those outputs.

Productivity is a concept that is closely related to efficiency. It is concerned with the ratio of an input (or aggregation of inputs) to an output (or aggregation of outputs). Productivity measures take no account of whether the observed variations in output can be attributed solely to the entity under scrutiny. Productivity measures are therefore usually more simplistic than efficiency measures, and may be less useful as a diagnostic tool. There have been

considerable efforts to apply the methods of efficiency analysis (regression analysis, stochastic frontier analysis and data envelopment analysis) to infer comparative efficiency of various aspects of health systems. Such techniques seek to build production functions by explaining justifiable variations in performance between the entities under scrutiny, and to characterize the unexplained residual as “inefficiency” (Street & Hakinnen, 2009). Although conceptually attractive, there are few examples of such methods being used in earnest by decision makers.

Comparative efficiency indicators are constructed using health care system data from the various stages of production processes, such as conversion of hospital inputs (capital, manpower, pharmaceuticals, devices, etc.) into valued outcomes, in the form of health improvement. The precise types of data used depend on availability and the scope of the entity under scrutiny. Examples of data from the various stages include:

- costs: sometimes disaggregated into categories, with or without overheads allocated;
- physical inputs: such as measures of labour (staff employed, by category) or capital (e.g. hospital beds);
- activities: such as procedures undertaken, days of care provided, diagnostic tests ordered, community visits made;
- physical outputs: such as episodes of care or patients cared for;
- outcomes: such as quality-adjusted life-years (QALYs), avoidable deaths, or other health status data.

Almost all efficiency indicators are constructed as a ratio of one of these measures (for example costs) to another (such as the related activities), offering an indication of the extent to which resources have been used efficiently along some or all of the production pathway.

There are various challenges involved in constructing meaningful efficiency indicators. A fundamental analytic challenge is to control adequately for the range of environmental factors, policy constraints, population characteristics and other factors that may be important constraints on productivity improvement. Even after undertaking some adjustment for such constraints, it is important to exercise caution in attributing all of the observed variation to variations in efficiency. Furthermore, the assignment of inputs and associated costs to specific health system activities can be problematic from an accounting perspective, often relying on arbitrary accounting rules or other questionable assignments. In summary, it may be difficult to ensure that the output being captured is directly and fully dependent on the inputs included in the measurement. Table 7 presents a sample of efficiency indicators commonly used by policy-makers,

Table 7. Sample of efficiency indicators

Indicator	What is it?	What are the assumptions and what does it ignore?
Cost-effectiveness of certain intervention	Cost per QALY	Assumes average costs of providing intervention do not change with scale; major data constraints.
Emergency department (ED) visits that could have been seen in less invasive settings	The proportion of ED visits that could have been seen in a different, less costly setting	Ignores quality of care. Depends on definitions.
Average length of stay	The number of days per hospital inpatient stay	Cases are identical, both in terms of outcomes and in terms of intensity.
Unit costs	Estimates of costs	Assume uniform treatment, uniform accounting methods, ignore quality.
Case-mix adjusted cost per episode of care	The average costs for treating a certain type of condition	Cases are identical, both in terms of outcomes and in terms of intensity; assumes uniform treatment, uniform accounting methods.
Duplicate medical tests	The number of tests that are done more than once for the same patient	Assumes any duplicate test is an inefficiency regardless of situation.
Share of total expenditures spent on administration	The percentage of total health expenditures dedicated to administration	Assumes that greater share of admin expenditure is inefficient without accounting for scale. Highly dependent on accounting methods used.
Labour hours per episode of care	The number of hours per case-mix adjusted episode of care	Assumes patients require the same intensity of care; difficult to measure accurately across a large sample; affected by health system design as well as efficiency.
Share of health worker hours spent treating patients	The percentage of health worker hours spent treating patients	Assumes patients require the same intensity of care; difficult to measure accurately across a large sample; assumes time not spent with patients is unproductive.

Table 7. Sample of efficiency indicators (continued)

Indicator	What is it?	What are the assumptions and what does it ignore?
Disease costs	The average cost per case of treating a certain disease	Can be difficult to calculate without linking patient data across providers. Assumes uniform case-mix. Highly dependent on accounting methods used.
Effective coverage	The share of actual health gains achieved relative to maximum potential health gains for an intervention	Difficult to measure need and quality.

Source: Cylus & Smith, forthcoming.

highlighting important assumptions that these indicators make. While efforts should be undertaken to improve accounting and data collection efforts, there is a limit to what can be achieved analytically in this domain, and policy-makers should make themselves familiar with the limitations of any indicators used.

4 Future directions and conclusions

If undertaken carefully, health system performance comparison offers a powerful resource for identifying weaknesses and suggesting relevant reforms. The progress that has been achieved is impressive, both in the scope of areas for which comparable international data on health are now available and in the degree to which comparability has been improved. However, the science of international comparison is at a developmental stage. Policy-makers therefore need to be made aware of both the strengths and limitations of health system comparison.

There are various ongoing initiatives and developments that have the potential to benefit further international comparisons. One very large area of development is that of information and communication technologies (ICT), often described within the EU context in particular as “e-health”. This area has the potential to greatly improve data collected at the system level. Moreover, as increasing numbers of people seek health care outside their own country, there is a growing incentive for better comparability at the international level (Busse et al., 2011b). However, certain challenges remain in securing comparability across conceptualizations, definitions and indicators in the international setting (Table 8).

Table 8. Key resources and challenges

Dimension	Challenges for international comparison	Way forward
Population health	<p>Many measures fail to distinguish the contribution of the health system.</p> <p>Many countries' mortality data are lacking in timeliness and accuracy.</p> <p>Problems of comparability among countries and over time, reflecting changes in and differences between international and national coding rules.</p> <p>Large gaps in availability of evidence on the effectiveness of treatments reducing mortality.</p>	<p>The development of Electronic Health Records (EHRs) and the development of multi-language software to assist with coding to deal with the problems in timeliness and accuracy of mortality data.</p> <p>The development of internationally comparable clinical databases that provide risk-adjusted information on individual outcomes of treatment that can help disentangle attribution.</p> <p>Greater use of tracer conditions which enable a better understanding of everyday experience of those in need of care.</p>
Health service outcomes	<p>Lack of well-defined boundaries of "health systems" and "health services".</p> <p>Gaps in understanding of the relationships between measurements on the micro-, meso- and macro-levels of the health system.</p> <p>Limited set of dimensions captured by outcome measures with a marked lack of measures on disabilities or discomfort.</p> <p>Intermediate process measures show potential for complementing outcome indicators but when used on their own may not be as meaningful to stakeholders.</p> <p>Lack of available, good-quality and comparative data at the patient level.</p> <p>International comparison is complicated by different organizational settings and reporting conventions across systems.</p>	<p>Creating more registry data, which identifies individual patients and traces them through the care process.</p> <p>Focus on a small number of indicators which could give an overall comparison of the quality of health systems, such as the HCQI project taken forward by the OECD and co-financed by the EU.</p> <p>Find ways to measure outcomes that are not defined in terms of cure, which are important for the measurement of chronic disease and long-term care.</p> <p>Find ways to assess health systems based on how well they perform from the perspective of the people that they are intended to benefit.</p>

Table 8. Key resources and challenges (continued)

<p>Equity</p> <p>Lack of existing datasets which provide a longitudinal perspective.</p> <p>Limited evidence has been recorded on how sensitive inequalities are to the inclusion of environmental effects.</p> <p>Limited understanding of the factors explaining the health production process and sources of inequalities, including the role of mental conditions along with cognitive biases in measuring self-reported health.</p> <p>Wide use of self-reported measures of health status given their availability in harmonized datasets, which allow international comparisons, but have important limitations which are aggravated for the lack of measures of calibration available in some datasets but not in others.</p> <p>Inadequate identification of what stands behind measures of socioeconomic position, namely different income sources and measures of wealth and social environmental controls which differ across the life-cycle.</p>	<p>Better collection of indicators on determinants of health.</p> <p>Invest in well-designed evaluative studies of major interventions to reduce health inequalities and equity to access in health care, taking advantage of natural experiments (changes in employment opportunities, housing provision or cigarette pricing).</p> <p>Put processes in place to ensure the availability and comparability of data, as well as harmonization of definition and collection instruments.</p> <p>Invest in data linkages to allow desegregation by socioeconomic status and better monitoring of health inequalities across countries.</p>
<p>Fairness in financing</p> <p>The construction of indicators depends crucially on adequately defining and measuring households' true capacity to pay for health care, which is not straightforward.</p> <p>Indicators only account for the short-term impacts of financial hardship and ignore the influence of coping strategies (such as selling assets or borrowing money to pay for care), with their longer-term consequences for household welfare.</p> <p>Indicators do not take into consideration the consequences of lost income due to illness for the measured degree of financial protection.</p> <p>Indicators do not give information about (and are likely to be affected by) the extent of financial barriers to access to health care.</p>	<p>Invest in data collection of household income and spending patterns.</p> <p>Research to develop metrics better suited to capture the intertemporal financial consequences of illness arising from coping strategies.</p> <p>Search for practical alternatives to account for the effect of financial barriers to health care access in financial protection analyses.</p>

<p>Responsiveness</p>	<p>Lack of conceptual clarity as to what constitutes responsiveness. Lack of clarity creates confusion as to what is measured and how. Surveys on satisfaction are very sensitive to question wording, sampling and demographic factors.</p> <p>The production process underlying health systems is intrinsically complex and poorly understood. Most measures make simplifying assumptions that may sometimes result in misleading data. Outputs are generally multidimensional and therefore preference weights are needed if they are aggregated into a single measure of attainment. The choice of such weights is intrinsically political and contentious. A fundamental challenge in developing an efficiency measure is ensuring that the output that is being captured is directly and fully dependent on the inputs that are included in the measurement. Environmental factors, policy constraints, population characteristics and other factors may be largely responsible for determining health outcomes, yet it is difficult to incorporate all possible determinants appropriately into an efficiency assessment. From an accounting perspective, the assignment of inputs and associated costs to specific health system activities is fundamentally problematic, often relying on arbitrary accounting rules or other questionable assignments. Although researchers have developed indicators that seek to measure full production processes, these measures are often not the most informative for policy-makers looking to identify and address inefficiencies. Many outputs are the results of years of health system endeavour, and cannot be attributed to inputs in a single period.</p>
	<p>Agree on a working international definition of responsiveness on which to base measures. Incorporate tools that allow for some correction of the bias produced by differences in experience or wider cultural differences.</p> <p>Improve data collection on key inputs which affect efficiency, such as staff and technology, where international data is underdeveloped. Research to find suitable metrics that measure organizational factors and administrative structures, which influence inputs and outputs. Improve clarification on the type of efficiency being measured by different indicators. Improve the conceptualization of the production process in order to better harmonize data collection efforts. Improve collection of high-quality comparable data on outputs, inputs and environmental factors necessary for risk adjustments. Invest in research to refine methodologies for whole-system efficiency measurement. Find a balance between whole-system measures and more fragmented efficiency measures. More consideration of how indicators take static and dynamic elements of inputs and outputs into account.</p>

Source: Fahy et al., forthcoming.

A particularly European issue, relevant to all domains of performance, is that although international comparisons are historically made between whole countries, in most European countries health systems are now organized primarily at a regional level. Therefore, comparisons at national level, without some regional breakdown, are inherently limited in their utility, as they may conceal wide internal variation. Some progress has already been made in developing and validating regional-level comparisons.

An important further consideration is that many of the indicators used for international comparison contain implicit value judgements that should be subjected to careful scrutiny. For example, concepts such as health outcomes, disability weights, responsiveness and equity assume a certain set of values as to what constitute the objectives of the health system, and what their relative importance is. Policy-makers at the very least need to be aware that certain value judgements have been made in how indicators are selected, measured and presented.

The presentation of comparisons has hitherto not been especially helpful for policy-makers. Neither the bald presentation of league tables nor a detailed narrative of caveats is well suited to securing appropriate policy responses. Two types of risk arise from poor presentation of comparisons: uncritical acceptance of results and potentially costly and inappropriate reforms of the health system; or rejection of the comparisons as inadequate, and a consequent lost opportunity to reform. In either case, the key issue is the need to focus on the policy-maker's action, and to ensure that it is well-informed, acknowledges the inevitable uncertainty, and is proportionate. For this to be achieved, it will usually be necessary to present indicators of health system environmental factors, functions and capacity alongside performance measures. These will assist in explaining the reported performance, and suggesting policy responses.

The key requirements necessary to create comparable indicators that address the needs of policy-makers are likely to be: appropriate methods of summarizing complex information; a narrative that picks out the key issues and uncertainties; a diagnosis of why the reported variations are arising; and the implications for policy action. It is nevertheless important to note that the comparisons might inform but should never be the sole criteria for recommending policy action. National policies, values and priorities should always be the starting point for policy action.

References

- Allin, S, Hernández-Quevedo C, Masseria C (2009). Measuring equity of access to health care. In: Smith PC et al., eds. *Performance measurement for health system improvement*. Cambridge, Cambridge University Press.
- Aday LA et al. (2004). *Evaluating the healthcare system: effectiveness, efficiency, and equity*. Chicago, Health Administration Press.
- Arah, OA et al. (2006). A conceptual framework for the OECD Health Care Quality Indicators Project. *International Journal for Quality in Health Care*, 17:5–13.
- Atun R, Mendabde N (2008). Health systems and systems thinking. In: Coker R, Atun R, McKee M (eds) *Health systems and the challenge of communicable diseases: experiences from Europe and Latin America*. European Observatory on Health Systems and Policies Series. Copenhagen, WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies.
- Blendon R et al. (1990). Satisfaction with health systems in ten nations. *Health Affairs*, 9(2):185–192.
- Busse R (forthcoming). Responsiveness. In: Smith PC et al., eds. *Health system performance comparison: an agenda for policy, information and research*.
- Busse R et al. (2011a). Being responsive to citizens' expectations? The role of health services in responsiveness and satisfaction. In: McKee M, Figueras J, Saltman RB, eds. *Health systems: health, wealth, society and well-being*. Oxford, Oxford University Press.
- Busse R et al. (2011b). *Diagnosis-related groups in Europe*. Maidenhead, Open University Press.
- Commonwealth Fund (2006). *Framework for a high-performance health system for the United States*. New York, The Commonwealth Fund.
- Coulter A, Cleary PD (2001). Patients' experiences with hospital care in five countries. *Health Affairs* 20(3):43–53
- Cylus J, Smith PC (forthcoming). Efficiency. In: Smith PC et al., eds. *Health system performance comparison: an agenda for policy, information and research*.
- De Silva A (2000). *A framework for measuring responsiveness*. Geneva, WHO (GPE Discussion Paper 32) (<http://www.who.int/responsiveness/papers/paper32.pdf>, accessed 13 February 2010).
- European Commission (1996). *Eurobarometer 44.3*. Brussels, European Commission.

- European Commission (1998). *Eurobarometer 49*. Brussels, European Commission.
- European Commission (1999). *Eurobarometer 50.1*. Brussels, European Commission.
- European Commission (2000). *Eurobarometer 52.1*. Brussels, European Commission.
- European Commission (2002). *Eurobarometer 57.2*. Brussels, European Commission.
- Fahy N., Klazinga, N., Lafortune, G. & Stein, C. (forthcoming). Commentary on international health system data. In: Smith PC et al., eds. *Health system performance comparison: an agenda for policy, information and research*.
- Grol R et al. (2000). Patients in Europe evaluate general practice care: an international comparison, *British Journal of General Practice*, 50:882–887.
- Hernández-Quevedo C, Papanicolas I, Machenback J (forthcoming). Equity. In: Smith PC et al., eds. *Health system performance comparison: an agenda for policy, information and research*.
- Hurst J, Jee-Hughes M (2001). *Performance measurement and performance management in OECD health systems*. Paris, Organisation for Economic Co-operation and Development (OECD Labour Market and Social Policy Occasional Papers No. 47).
- IHP (2008). Monitoring performance and evaluating progress in the scale-up for better health: a proposed common framework. Document prepared by the monitoring and evaluating working group of the International Health Partnership and Related Initiatives (IHP+) Led by the WHO and the World Bank.
- Jee M and Or Z (1999). *Health outcomes in OECD countries: a framework of health indicators for outcome-oriented policymaking*. Paris, Organisation for Economic Co-operation and Development (OECD Labour Market and Social Policy Occasional Papers No. 36).
- Jenkinson C, Coulter A, Bruster S (2002). The Picker Patient Experience Questionnaire: development and validation using data from in-patient surveys in five countries. *International Journal for Quality in Health Care*, 14:353–358.
- Jones, AM, A López-Nicolás (2004). Measurement and explanation of socioeconomic inequality in health with longitudinal data. *Health Economics*, 13(10):1015–1030.
- Karanikolos M et al. (forthcoming). Health outcomes: measuring population health. In Smith PC et al., eds. *Health system performance comparison: an agenda for policy, information and research*.

- Kelley E, Hurst J (2006). *Health Care Quality Indicators Project: conceptual framework paper*. Paris, Organisation for Economic Co-operation and Development (OECD Health Working Papers No. 23).
- Klassen A et al. (2009). Performance measurement and improvement frameworks in health, education and social services systems: a systematic review. *International Journal for Quality in Health Care*, 22:44–69.
- Klazinga N, Li L (forthcoming). Health service outcomes. In: Smith PC et al., eds. *Health system performance comparison: an agenda for policy, information and research*.
- Moreno-Serra R, Millett C, Smith PC (2011). Towards improved measurement of financial protection in health. *PLoS Medicine*, 8(9):e1001087.
- Moreno-Serra R, Thompson S, Xu K (forthcoming). Financial protection. In: Smith PC et al., eds. *Health system performance comparison: an agenda for policy, information and research*.
- Murray CJL, Frenk J (2000). A framework for assessing the performance of health systems. *Bulletin of the World Health Organization*, 78:717–730.
- Neely A (forthcoming). Benchmarking. In: Smith PC et al., eds. *Health system performance comparison: an agenda for policy, information and research*.
- Nolte E, McKee M (2004). *Does healthcare save lives? Avoidable mortality revisited*. London, The Nuffield Trust.
- Nolte E, Bain C, McKee M (2009). Population health. In: Smith PC et al., eds. *Performance measurement for health systems improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press.
- O'Donnell O et al. (2008). *Analyzing health equity using household survey data*. Washington, DC, World Bank Institute, World Bank.
- OECD (2010). *Improving value in health care: measuring quality*. Paris, Organisation for Economic Co-operation and Development.
<http://www.oecd.org/health/hcqi>.
- Papanicolas, I., Smith, PC. (forthcoming) Reviewing International Frameworks for Health System Performance Assessment. In: Smith PC et al., eds. *Health system performance comparison: an agenda for policy, information and research*.
- Rice N, Robone S, Smith PC (2010). International comparison of public sector performance: the use of anchoring vignettes to adjust self-reported data. *Evaluation*, 16(1):81–101.

Roberts MJ et al. (2008). *Getting health reform right: a guide to improving performance and equity*. Oxford, Oxford University Press.

Schoen C et al. (2007). Toward higher-performance health systems: adults' health care experiences in seven countries, 2007. *Health Affairs*, 26(6): w717–w734.

Sicotte C et al. (1998). A conceptual framework for the analysis of health care organizations' performance. *Health Services Management Research*, 11:24–48.

Smith PC et al. (eds.) (2009). *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press.

Street A, Hakinen U (2009). Measuring productivity. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press.

Valentine N et al. (2009). Health systems responsiveness: a measure of the acceptability of health care processes and systems from the user's perspective. In: Smith PC et al., eds. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge, Cambridge University Press.

Wagstaff, A. van Doorslaer, E. van Der Burg, H. Calonge, S. Christiansen, T. Citoni, G. Gerdtham, UG. Gerfin, M. Gross, L. Häkinnen, U. Johnson, P. John, J. Klavus, J. Lachaud, C. Lauritsen, J. Leu, R. Nolan, B. Perán, E. Pereira, J. Propper, C. Puffer, F. Rochaix, L. Rodríguez, M. Schellhorn, M. Winkelhake, O. et al (1999). Equity in the finance of health care: some further international comparisons. *Journal of Health Economics*, 18(3): 263–290.

Wagstaff A, van Doorslaer E, Watanabe N (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. *Journal of Econometrics*, 112:207–223.

WHO (2000). *World health report 2000. Health systems: improving performance*. Geneva, World Health Organization.

WHO (2010). *World health report 2010. Health systems financing: the path to universal coverage*. Geneva, World Health Organization.

Joint policy summaries

1. Addressing financial sustainability in health systems
Sarah Thomson, Tom Foubister, Josep Figueras, Joseph Kutzin, Govin Permanand, Lucie Bryndová
2. Assessing future health workforce needs
Gilles Dussault, James Buchan, Walter Sermeus, Zilvinas Padaiga
3. Using audit and feedback to health professionals to improve the quality and safety of health care
Signe Agnes Flottorp, Gro Jamtvedt, Bernhard Gibis, Martin McKee

The **Health Evidence Network** (HEN) of the WHO Regional Office for Europe is a trustworthy source of evidence for policy-makers in the 53 Member States in the WHO European Region. HEN provides timely answers to questions on policy issues in public health, health care and health systems through evidence-based reports or policy briefs, summaries or notes, and easy access to evidence and information from a number of web sites, databases and documents on its web site (<http://www.euro.who.int/en/what-we-do/data-and-evidence/health-evidence-network-hen>).

The **European Observatory on Health Systems and Policies** is a partnership that supports and promotes evidence-based health policy-making through comprehensive and rigorous analysis of health systems in the European Region. It brings together a wide range of policy-makers, academics and practitioners to analyse trends in health reform, drawing on experience from across Europe to illuminate policy issues. The Observatory's products are available on its web site (<http://www.healthobservatory.eu>).