Performance management in health systems and services: Studies on its development and use at international, national/jurisdictional, and hospital levels
Veillard, J.H.M.

Citation for published version (APA):
Veillard, J. H. M. (2012). Performance management in health systems and services: Studies on its development and use at international, national/jurisdictional, and hospital levels

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 3
International health system comparisons: From measurement challenge to management tool
Abstract

Objectives. To review current experiences in using international health system performance comparisons for performance management purposes and analyse methodological issues and challenges in carrying out international health system performance comparisons.

Methods. We used a purposive review of the literature and a case study of the OECD experience with the Health Care Quality Indicators Project.

Results. We observed that motivations for international health system performance comparisons included an increasing need by policy-makers to be accountable to the public, the desire of policy-makers for better strategic planning to meet desired outcomes, and the mutual benefit of benchmarking for health system re-engineering. We also noted that previous initiatives varied widely in scope. Some conducted comparisons between national health systems using broad indicators while others conducted disease or sector-specific comparisons cross-nationally. Using the OECD-HCQI project as an example, we compiled the methodological barriers to international comparisons into six key issues: specifying indicators using internationally standardized definitions; controlling for differences in population structures across countries; adjusting for differences in information systems’ ability to track individual patients; controlling variability of data sources; identifying nationally representative data; and determining retrospective completeness of the time series. Finally, we identified a number of innovative approaches to using health system performance comparisons for performance management purposes and discussed potential policy implications.

Conclusions. International health system performance comparisons are increasingly used for performance management purposes, which calls for further research in defining the characteristics of practice benchmarking systems supporting policy learning and performance improvement.
Introduction

International comparisons of health system performance as provided by multilateral organizations such as the World Health Organization (WHO) and the Organisation for Economic Co-operation and Development (OECD) generate much interest. The provision of comparative data presents vast methodological challenges but offers considerable potential for cross-country learning. Policy-makers are looking for examples, benchmarks, and solutions to address the pressures imposed by the epidemiological, economic, societal, and technological demands on all European health care systems.

The use of international performance indicators to assess national economies and public domains such as education, transport, and environment has paved the way for their acceptance in the health care field. Dating back to the 1930s [1], studies on health insurance programmes in Western Europe show that international comparisons of health systems were used as a means to guide policy processes [2]. Several decades ago, such international assessments focused mainly on structural characteristics (e.g. numbers of physicians, nurses, and hospitals) and a few specific outcome parameters (e.g. perinatal mortality, under-five mortality, maternal death, incidence and prevalence of infectious diseases, average life expectancy at birth). In the European region these parameters were complemented by the work done around avoidable deaths [3] and the release of the first atlas of avoidable deaths in the European Union [4,5], thus introducing attempts to assess the contribution of health care to the overall health of populations. Coupled with data on health expenditures [6,7], these produced the first picture on the performance of national health systems in relation to resources used.

The publication of WHO’s World Health Report 2000 and the OECD’s Health at a Glance 2001 received (and continues to receive) much attention. The World Health Report 2000 was based on a generic conceptual performance framework and ranked Member States in a league table. Despite many criticisms (see Box 1), the report placed international health system performance on the political agenda; raised awareness about performance issues; and resulted in many initiatives to improve the perceived health situation in different countries. The latest version of Health at a Glance [8] contains a comprehensive array of performance indicators without attempting to group the findings in league tables. This has elicited a more nuanced reaction from participating countries. The OECD experience underscores the fact that comparative data help primarily by raising questions about the performance of health care systems rather than explaining why one country performs better than another.
Box 1. Debates around the World Health Report 2000

The World Health Report in 2000 was subject to a great deal of controversy. The following points summarise the key controversies pertaining to the political, technical, and methodological aspects of the report [9]:

- **Underlying political philosophy**—political and ideological debates accused the report of being too medical-model based and criticized its failure to consider the importance of primary health care systems.

- **Face validity**—experts questioned the actual rankings of certain countries. For example, the United States ranking higher than Denmark in the responsiveness measure despite the latter having a system of universal health care coverage.

- **Coherence of performance measures**—the report was criticized for mainly focusing on health care systems instead of considering broader social and educational factors and not accounting for the lag between health interventions and their measurable impact.

- **Data availability**—one of the greatest areas of contention was the use of estimates instead of actual data.

- **Health levels and distribution**—critics questioned the use of specific measures such as Disability-Adjusted Life Expectancy and equality measures.

- **Responsiveness levels and distribution**—another major area of contention was the use of limited key informants for assessing responsiveness of health systems and not considering the political contexts that could impact this measure.

- **Fairness of financing**—critics disputed the definitions and methods used to assess the fairness of financing measures.

- **Estimating performance**—several debates questioned the “achievement of performance in health system” concept used in the report.

- **Composite index**—the use of a composite index (especially the weighting methods used in the report) to measure health systems was heavily questioned.

- **Use of evidence**—many criticized the report for using a narrow evidence base.

Despite these debates, the World Health Report 2000 fostered the importance of health systems; its publication emphasized the need for health stewardship within national governments and played a significant role in raising the profile of accountability for health on political agendas. Following the release of the report, numerous countries such as Kyrgyzstan requested technical support from the WHO to revise their national health system policies and strategies. Furthermore, it created an impetus for further cross-national discussions around the importance of developing comparable data standards that can be utilized towards strengthening health system performance in countries.
This chapter discusses some of the main issues involved in international health system comparisons. The first two sections will examine the rationale (why) and the scope (what) of cross-national health system performance assessments, emphasizing various functions of comparisons (accountability, strategy development, and learning) and the scope of such efforts (whole systems, specific services, specific diseases, sub-national approaches). Using the OECD's HCQI project as an example, the third section deals with outstanding methodological issues and challenges (how) such as population variations, data standardization problems, differences in coding practices, and definitional issues that arise during international comparisons. The final section addresses the question of how countries can move from measurement to management by illustrating new initiatives that ensure cross-system data comparisons become an integral part of health system performance management and decision-making processes.

**Increased interest in international health system comparisons**

Several reasons underlie the increased interest in international health system comparisons. Firstly, in resource-scarce environments, the public and the media are increasingly holding policy-makers accountable. International data therefore play a key role in the accountability agenda, which enables countries to demonstrate that their performance on specific items is equivalent to (or better than) that reported in other countries. Various surveys indicate that accountability can be a generic function of governments towards their citizens but users’ negative experiences of health systems can also increase the pressure for governments to seek out best practices and policy lessons from other settings [10]. Additionally, the issue of patient responsiveness has recently gained momentum at the European level and could impact on future policy agendas in several countries. Furthermore, patient mobility adds an additional layer of public pressure on governments as borders become more porous in the European region [11,12].

Secondly, performance information from international comparisons, along with trend data and careful policy analysis, can form the input for national strategy development [13]. Following the application of balanced scorecards and strategy maps in the private finance industry [14,15], a growing number of countries are in the process of developing frameworks to assess their health systems through national performance reports and strategy development. Examples of such reports are found in the United States [16,17], Ontario, Canada [18], and the Netherlands [19]. Similarly, the use of balanced scorecards has impacted the establishment of information systems and the management and delivery of health care services at national and sub-system levels [20,21]. International benchmarking data can thus help in formulating the
national policy programme. However, it is necessary to use a cautious approach when using comparative data for strategy development purposes since hidden political agendas and selective perception can distort the performance evidence [22].

Thirdly, other systems gain opportunities to learn from and emulate the efforts of effective restructuring successes based on performance data from health systems such as the Veterans Health Administration in the United States [23]. Thus mutual learning constitutes the third function of international health system comparisons. As data become more robust it becomes feasible to analyse the factors contributing to better performance—this constitutes an important part of the still limited evidence-based knowledge on health system engineering. The value of sharing similar challenges and experiences is greatly enhanced when governments identify peer groups for comparison. For example, the Nordic Council of Ministers is involved in efforts to compare the quality of care among their countries—Denmark, Finland, Iceland, Norway, and Sweden. The results of the study are intended for use in monitoring and evaluating health services while providing a forum for sharing learning experiences amongst participating countries [24].

In summary, accountability and strategy development are currently the major functions driving governments to engage in international health system comparisons. However, mutual learning is gaining further interest with the increasing scientific robustness of knowledge created through health systems research.

**Scope of international health system comparisons**

The scope of international health system comparisons varies by country, type of established health information system, and availability of resources. The first stage in setting up an international comparison comprises the development or identification of a conceptual framework against which the utility and validity of a set of indicators can be assessed. International organizations have presented conceptual frameworks that aim to describe the underlying constructs, domains, and their mutual relations. For example, WHO and the OECD developed such frameworks for health system performance assessment to form the basis for the World Health Report 2000 and a frame for the HCQI project, respectively [25] (see Box 2).
Box 2. Standardization of performance concepts in international health system comparisons – WHO and OECD conceptual frameworks

**WHO health system performance measurement:** The tiers chosen by WHO in conceptualizing performance are multi-dimensional, reflecting those considered as the main goals of a health system: improvement of population health, responsiveness to population expectations, and fairness in financial contribution across the population. Figure 1 provides a summary of the main features of this framework. Additionally, in order to provide a relevant policy context for the performance of a health system, four main functions were identified: stewardship, financing, service provision, and resource generation.

**Figure 1. WHO Health System Performance Framework**

<table>
<thead>
<tr>
<th>Components for assessment Goals</th>
<th>Average level</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health improvement</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Responsiveness to expectations</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fairness in financial contribution</td>
<td>———</td>
<td>✓</td>
</tr>
</tbody>
</table>

Source: Murray and Frenk, 2000 [26]

**Figure 2. Boundaries of Health systems in the WHO conceptual framework**

**OECD HCQI conceptual framework:** Similar to the conceptualization by WHO, the OECD also adopted a multi-dimensional approach. The framework below presents a visual summary of the dimensions of health care performance, including: quality, access, cost, efficiency, and equity. It also presents a picture of factors related to, but distinct from, health system performance, such as: health system design, policy, and context; non-health care determinants of health and overall levels of health. Finally, it highlights the particular dimensions of quality of care that are the focus of the HCQI project, namely effectiveness, safety, and responsiveness or patient experience.

Source: Murray and Evans, 2003 [27]
The design of a proper set of indicators within such frameworks necessitates the initial, unavoidable task of answering fundamental questions relating to the definition of health system performance, selection of measures, and interaction among the individual indicators. The set cannot be a random list of measures or a simple repository of information and is normally conceived as a system articulating information with a certain purpose—in the case of WHO and OECD, to inform the comparative performance of health systems. There is consensus that indicators selected to compare performance should (i) be scientifically solid; (ii) be politically relevant; (iii) be available across a sufficient number of countries; and (iv) allow for sustainable and feasible data collection across time [26,27].

These frameworks developed by international organizations encompass structures used in several existing national performance reports [29] and, as Arah et al. (2003) noted [30], contain many similar dimensions and perspectives. For a
classification of the on-going health system comparisons one can also look at whole system, multilateral, bilateral, disease, sector- or domain-specific approaches. Table 1 provides a broad categorization of different types of international comparisons of health systems. Some are undertaken on a regular, systematic basis (e.g. OECD HCQI project); others were one-time comparisons (e.g. between the United Kingdom’s NHS and California’s Kaiser Permanente). Although the list is by no means comprehensive, many of these endeavours seek to overcome epidemiological, economic, or geopolitical considerations by identifying specific components of the health system and measuring performance on those factors.

Table 1. General classification of health system comparisons

<table>
<thead>
<tr>
<th>Type of initiative</th>
<th>Systems/factors involved</th>
<th>Selected examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire health system</td>
<td>Broad comparisons of overall health systems</td>
<td>World Health Report 2000 [31]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Commonwealth Fund studies comparing high-performing health systems of the United States, United Kingdom, Canada, Germany, Australia and New Zealand [32]</td>
</tr>
<tr>
<td>Multilateral</td>
<td>Comparisons between national or sub-national health systems</td>
<td>Commonwealth Fund study on health system comparisons of six countries that measure various dimensions of health-care systems including quality, access, equity, efficiency and healthy lives [32]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>European Commission-funded project: Indicateurs de Santé des Régions Européennes (ISARE) covers 283 health regions in 24 European countries [33]</td>
</tr>
<tr>
<td>Bilateral</td>
<td>Comparisons between national health systems; national health systems and provincial regional health systems; or national health systems and health-care organisations</td>
<td>Comparison of health system in Canadian province of Ontario and health system in the Netherlands [34]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comparison of the United Kingdom’s NHS and California’s Kaiser Permanente in the United States [35]</td>
</tr>
<tr>
<td>Disease-specific</td>
<td>Comparisons of specific health conditions across countries/regions</td>
<td>Joint WHO/European Commission project: Benchmarking Regional Health Management (Ben RHM) covering 19 regions in 15 European countries and tracking 3 conditions – diabetes, breast cancer and measles [36,37]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nordic Council of Ministers’ comparisons of specific disease conditions in Denmark, Finland, Iceland, Norway and Sweden [24]</td>
</tr>
<tr>
<td>Sector-specific</td>
<td>Comparisons of segments of the health care system e.g. primary care</td>
<td>Comparison of primary care systems for 18 OECD countries from 1970-1998 [38]</td>
</tr>
<tr>
<td>Domain-based</td>
<td>Comparisons among components of the health-care system e.g. waiting times, patient experiences</td>
<td>OECD HCQI project involving 30 countries [25]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Commonwealth Fund study on patient experiences in 7 countries [39]</td>
</tr>
</tbody>
</table>
As noted earlier, initiatives such as those undertaken by the WHO and OECD assess a broader set of health measures than those studied in traditional comparisons of health systems (e.g. health expenditures among countries; indicators such as life expectancy). Taken a step further, countries and international agencies are increasingly implementing sub-level comparisons, especially at the European Union level. For example, Ben RHM and ISARE are two European Commission funded projects that identified European regions with some common features in their political, socio-demographic, and epidemiological development and initiated benchmarking efforts to determine the structural, functional, and quality differences of health services within the selected countries. Experiences from these projects show that smaller countries often prefer comparative efforts in which they are evaluated against regions, rather than the entire national health system, of bigger countries [33]. Furthermore, sub-level comparisons enabled networking opportunities among health experts and fostered mutual learning experiences [36].

A major reform of their health system provides a unique opportunity for countries to undertake comparative studies, allowing related policy and performance changes to be monitored. In 2006, following such a restructuring, the Netherlands initiated a comparative study of their health sector and that of Ontario, Canada, which had undergone reforms during a similar time period. Both Ontario and the Netherlands invested in the development of reliable health system performance assessment frameworks. The study mapped various dimensions of these and compared each of the systems. Conceptual and contextual problems prevent the two systems from being completely comparable but they still provide a starting point for such benchmarking efforts and highlight the range of issues involved in international comparisons [34].

Some researchers have attempted to overcome larger methodological barriers of cross-country assessments by examining specific components of health systems. For example, a controversial study by Feachem et al. in 2002 compared performance factors such as access and responsiveness in the British NHS to the California branch of Kaiser Permanente in the United States [35]. The authors concluded that Kaiser Permanente performed better and had a better integrated and managed system than the NHS, despite similar costs. The study was heavily criticized for flaws in both its methodology and its assumptions [40] and illustrates that, while individual components of health systems can be compared, it is imperative that such exercises are approached with caution.

This discussion of the various comparative projects is far from complete but illustrates the type of work currently being implemented. In addition, it should be mentioned that major developments are underway to increase the potential of international comparisons in health care at the level of both international research and of cross-system databases. At the research level, studies in areas such as cancer care, cardiovascular diseases, and diabetes have largely increased the availability of
international comparative data. Research projects funded by the EC (e.g. Ben RHM, ISARE) are good examples of this type of work currently being implemented. The field of health systems analysis has also expanded and various targeted research groups have been established over the past decade.

Apart from these research processes, expert working groups in international organizations are leading efforts to increase data comparability among countries. Along with the WHO work on classification of diseases (ICD-9, ICD-10, ICD-11) [41] and the OECD’s focus on comparing national health accounts and health financing data (OECD System of Health Accounts), there is active collaboration among WHO, OECD, and the European Union [42] to improve the comparability of national data systems.

By contrast, several transition countries in the European region are still establishing their health information systems and therefore comparative studies occur on a limited basis. However, as a first step, a number of countries are involved in the Health Metrics Network (http://www.who.int/healthmetrics) which is hosted by WHO and enables them to overcome problems of data availability and improve the quality and reliability of their information systems. Although some transition countries lack optimal quality control measures, many are increasing investments in efforts to align their health systems with international standards. For example, WHO recently led initiatives by which Armenia and Kyrgyzstan developed performance assessment frameworks to aid them in strengthening their health sector. In the long run, such endeavours will lead to benchmarking among comparable countries in the region and highlight areas for improvement in health system performance.

As seen in this section, international health system performance comparisons have a broad scope. Such assessments depend largely on project aims, policy opportunities, and availability of resources and data. Each type of comparison—from multilateral to domain specific—serves an important function in drawing attention to a particular health system and possible ways to strengthen its performance.

Methodological issues in conducting international health system comparisons: lessons from the OECD experience

Initiatives to build relevant and meaningful indicators across different countries face numerous challenges. This section provides an overview of the operational and methodological issues involved in such efforts. The matters explored follow the experience within the OECD HCQI project, but can be generalized to comparative efforts in similar international health systems.

The OECD HCQI project started in 2002 with the objective of developing a set of health care quality indicators that can be reported reliably and regularly across 30 OECD countries. The purpose was to help raise questions for further investigation into differences in the quality of care across countries. The number of countries involved
in the HCQI project has recently expanded to include all European Union Member States, including non-OECD nations, following an agreement between the European Commission Directorate-General for Health and Consumers and the OECD.

The HCQI project has undergone several phases. The initial list of indicators consisted of 86 potential measures in five priority areas of care (patient safety; mental health care; health promotion, prevention, and primary care; cardiac care; and diabetes care). However, data availability proved to be a major hurdle [28].

There has been a two-pronged strategy to overcome this barrier: (i) initiate regular data collection of widely available indicators; and (ii) simultaneously work with countries to improve information systems and enhance the comparability of data indicators. At the current state of development, the regularly updated set covers health areas outlined in Table 2 [43]. In addition, fourteen measures for patient safety and two for mental health care have reached the last phase of piloting and it is envisioned that they will be included in the regular set for 2009 data collection. The indicator set includes both process and outcome measures since they provide different but complementary insights—information derived from process indicators is easier to translate into specific improvements; outcome indicators may be subject to multifactor causal attribution but are indispensable in aligning performance assessment with health system objectives. The key is to establish a balance between these two types of measures.

**Table 2. HCQI project indicators**

<table>
<thead>
<tr>
<th>Care type</th>
<th>Outcome</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>For acute conditions</strong></td>
<td>In-hospital AMI case-fatality rates</td>
<td>Waiting times for surgery after hip fracture Age 65+</td>
</tr>
<tr>
<td></td>
<td>In-hospital ischemic/hemorrhagic stroke case-fatality rates</td>
<td></td>
</tr>
<tr>
<td><strong>For cancer</strong></td>
<td>Survival rate for colorectal cancer</td>
<td>Mammography screening</td>
</tr>
<tr>
<td></td>
<td>Survival rate for breast cancer</td>
<td>Cervical cancer screening</td>
</tr>
<tr>
<td></td>
<td>Survival rate for cervical cancer</td>
<td></td>
</tr>
<tr>
<td><strong>For chronic conditions</strong></td>
<td>Hospital admission rate for asthma (people 18+)</td>
<td>Annual retina exam for diabetics</td>
</tr>
<tr>
<td></td>
<td>Asthma mortality rate (people aged 5-39)</td>
<td></td>
</tr>
<tr>
<td><strong>For prevention of</strong></td>
<td>Incidence of measles</td>
<td>Vaccination against measles</td>
</tr>
<tr>
<td>communicable diseases</td>
<td>Incidence of Pertussis</td>
<td>Vaccination against Pertussis (and diphtheria and tetanus)</td>
</tr>
<tr>
<td></td>
<td>Incidence of Hepatitis B</td>
<td>Vaccination against Hepatitis B</td>
</tr>
<tr>
<td></td>
<td>Vaccination against influenza (people 65+)</td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>Smoking rates</td>
<td></td>
</tr>
</tbody>
</table>
Within the HCQI project, indicators are considered ready for international comparisons once the agreed threshold of 10 countries can provide data from well-identified and stable databases according to agreed definitions (age group, codes, methods of identification). Indicators are added and deleted in order to ensure that the set remains responsive to changes in data availability or measurement quality. The tension between maintaining a stable set over time and the imperative to convey a concise message to policy-makers should be balanced while making decisions. Furthermore, there is a trade-off between implementing rigorous methodological approaches and including all countries in the calculations. A balance point is achieved when the methodology is strict enough to provide policy insights but flexible enough to allow participation by the maximum number of countries. Another compromise is to achieve homogeneous information systems without overburdening the countries that are required to comply with such constraints, especially those bearing the cost of adding new data items to their collection structures. The improvement of national health information systems can be considered a positive side effect of the involvement in international performance assessment initiatives but any changes must take account of existing structures.

The OECD HCQI project provides a rich empirical experience of dealing with complex methodological barriers. Several key issues that need to be considered when establishing and monitoring cross-country performance indicators are as follows:

1. Specifying indicators using internationally standardized definitions.
2. Controlling for differences in population structures across countries.
3. Adjusting for differences in information systems’ ability to track individual patients.
4. Controlling variability of data sources.
5. Identifying nationally representative data.
6. Determining retrospective completeness of the time series.

These are described in the following sub-sections together with suggestions to overcome them.

**Specifying indicators using internationally standardized definitions**

Standardization constitutes the best way to ensure data comparability across countries since it is applied across all stages of data production, storage, and report.

WHO leads the main initiative in this field through the WHO Family of International Classifications (WHO-FIC) programme, comprising three types of systems:

1. International Classification of Diseases (ICD)
2. International Classification of Health Interventions (ICHI)
3. International Classification of Functioning, Disability and Health (ICF)
The ICD is used to classify diseases and other health problems and has become the international standard diagnostic classification for epidemiological and health management purposes. ICD-10 is the latest version (an updated ICD-11 is currently under development). However, countries can find it difficult to update to new versions of ICD as its impact in shaping national information systems involves issues such as staff training, adapting to new definitions, and changes to funding schemes. For example, ICD-10 contains 12,640 codes while ICD-9 had only 6,969 codes. As a consequence, the use of different versions of ICD across countries is a real issue while attempting to identify indicators for international comparison.

In the absence of an internationally accepted system for reconciling ICD-9 and ICD-10, the HCQI project has opted to develop ad hoc validated crosswalks for the indicators relying on them. The first initiative comprises 14 patient safety indicators that are currently being tested for adoption in 2009. The International Methodology Consortium for Coded Health Information (IMECCHI) is an expert network that has worked with the HCQI project to develop and validate a manual for the calculation of these measures. Consideration of both ICD versions and the national adaptations of ICD-10 provides a solid basis for ‘translation’ and enhancing comparability across countries [44].

There are other outstanding issues concerning the calculation of indicators based on standardized codified databases. For instance, actions to address variation in documentation and coding practices across countries will entail some cultural changes that take time. However, participation in international initiatives has the beneficial effect of drawing attention to practices that might be regarded as adequate at the national level, but become less acceptable when compared to those in similar countries.

The current lack of international classification system for procedures is another relevant aspect, especially for the specification of process indicators. The ICHI covers a wide range of measures for curative and preventive purposes but is still in its beta trial version and entering extensive field trials before being submitted for endorsement by the governing bodies of WHO [41]. Despite encouraging progress, it may be several years before ICHI will be ready for adoption; therefore the HCQI project currently utilizes ICD-9-CM and ICD-10 to specify procedures.

Endorsed in 2001, the ICF seems promising. However, it is not yet used widely across countries and its specific applicability in defining outcome indicators needs to be explored further.

**Controlling for differences in population structures across countries**

A number of indicators can be affected by a country’s demographic structure. For example, survival or mortality rates are influenced by the age and gender structure of the population. This demographic composition has an impact on the epidemiology
of diseases and becomes a confounding factor that assessments need to adjust for. Age and sex standardization facilitates comparisons across countries by controlling for these differences in national populations.

When selecting a reference population it is important to decide whether to use the general population or one that is disease-specific (i.e. has the distribution of patients with the respective disease). As the incidence and prevalence of most diseases increases with age, disease-specific populations tend to weigh older population segments more heavily. A disease-specific reference population is therefore theoretically superior but is frequently not feasible as it requires the construction of a population for each disease. Many research projects overcome this problem by using general population weights. Another technique reduces distortion by removing the segment of the population that is less affected by the disease, truncating the sample to include only those above a certain age, e.g. 40 [45].

The HCQI project initially considered the 1980 OECD population structure for age-standardization calculations. This decision is now being revised because (i) the structure of this population is becoming outdated with the demographic aging trends in OECD societies, and (ii) the OECD has expanded from 24 to 30 countries and therefore the 1980 reference has limited validity. The transition to a 2005 OECD reference population is under assessment. The adoption of a truncated population is also being analysed, especially as countries such as Japan face a higher prevalence of myocardial infarction in the elderly group rather than the typical middle-age range.

There is a trade-off between updating the structure of the reference population and maintaining valid comparable data over time. Other international comparative projects face similar challenges caused by aging populations and incorporating new member countries, e.g. European Union’s development of the European Community Health Indicators Monitoring project or the European Health Interview Survey [46,47]. Steps should therefore be taken to ensure that the data remain valid and comparable over time.

**Adjusting for differences in information systems’ ability to track individual patients**

Indicators often take the form of rates in which the denominator is a specific group of patients—this cluster of indicators includes hospital fatality rates among patients with certain diagnoses or rates of specific procedures among chronically ill patients. Two interrelated issues affect the feasibility of these indicators: (i) the need to distinguish between different patients and repeated events affecting the same patient; and (ii) the necessity of detecting a patient’s contact at any level of care and across different institutions. However, national information systems do not have a uniform ability to identify patients and often the only data available are activity records that count each episode of care separately, even if the same patient was involved.
Hence, there is a clear need to harmonize calculations across countries to ensure data comparability; Mattke et al. in 2006 illustrate the effect of different bases of calculation on 30-day hospital fatality rates for myocardial infarction and stroke [28]. Currently, the most generally feasible approach is events-based calculations in which it can be reasonably argued that the validity of a specific indicator is not affected. However, a unique patient identifier is the most efficient tool for performing patient-based calculations and the OECD recently began encouraging member countries to establish these across their key health information systems.

**Controlling variability of data sources**

National information systems comprise a variety of data sources with substantial differences in their structure; the nature of data recorded; and the purpose for which they were conceived. Data systems have been shaped to serve monitoring functions within each country. Often, the purpose of such monitoring is not performance comparison or quality measurement but rather to support administrative activities such as budget distribution or system management (see Box 3 for a summary of main data sources and their general strengths and weaknesses). This means that a fair assessment of the available sources across countries and their suitability (on an indicator by indicator basis) will be required when building indicators for international comparison. For instance, process indicators such as vaccination or screening rates can be built from data from varying sources across countries but the nature of available data will vary with the structure of health service provisions in each system.

**Box 3. Sources of information available to assess quality of care across countries**

<table>
<thead>
<tr>
<th>Source</th>
<th>Weaknesses</th>
<th>Strengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Administrative data</td>
<td>Limited/no information on processes of care and physiologic measures of severity</td>
<td>Data availability improving Coding systems (international classifications of diseases) and practices are improving</td>
</tr>
<tr>
<td>Admission/discharge records</td>
<td>Limited/no information on timing (co-morbidities vs. onset or adverse events)</td>
<td>Large data sets optimize precision Comprehensiveness (all hospitals, all payers) avoids sampling/selection bias</td>
</tr>
<tr>
<td>Minimum set of data</td>
<td>Heterogeneous severity within some ICD codes</td>
<td>Data are used for other purposes and therefore subject to auditing and monitoring</td>
</tr>
<tr>
<td>Insurance-reimbursement</td>
<td>Accuracy depends on documentation and coding</td>
<td></td>
</tr>
<tr>
<td>DRGs accounting</td>
<td>Data are used for other purposes, subject to gaming</td>
<td></td>
</tr>
<tr>
<td>Prescription</td>
<td>Variation in how administrative data are collected and used, in particular DRG-based payment versus global budgeting versus service-based payment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time lag may limit usefulness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poor development outside the hospital setting</td>
<td></td>
</tr>
</tbody>
</table>
In some countries, prevention activities are organized in large-scale national programmes with routine databases that can be used for analysis. However, data in other countries are managed by each municipality and therefore registries are fragmented and not always accessible at the national level. In addition, registries for prevention activities often do not cover settings outside the health care system (e.g. work or school) and private organizations that provide this type of care can vary by country, complicating the retrieval of documented activity. In other cases, programmes are non-existent and services are provided on a demand basis. In all these situations, population surveys might be the most valid source of information.

The key question is whether data from so many different data sources (registries and population surveys) are comparable. As part of a methodological refinement, the HCQI project assessed the data comparability of surveys and programme registries for cancer screening indicators. Median rates of mammography and cervical cancer

<table>
<thead>
<tr>
<th>Source</th>
<th>Weaknesses</th>
<th>Strengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. National surveys</td>
<td>Self-reported (recall bias, lack of accuracy due to lay approach of those interviewed) Inability to identify and follow up subjects</td>
<td>Population based rather than patient based information, including individuals that health information systems cannot account for Can provide a basis for access and needs assessments</td>
</tr>
<tr>
<td>Health status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health services use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmaceutical consumption</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. National registries</td>
<td>When not mandatory, some eventual selection bias may deem them not representative Resource intensive to register the detailed specific features (e.g. adding cancer staging data to the diagnosis in the cancer registries) Not always linkable to other sources of information</td>
<td>Precise specific information</td>
</tr>
<tr>
<td>Cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic diseases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse events</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certain procedures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Medical records</td>
<td>Data retrieval is work intensive and therefore expensive, even with electronic records Difficult to sustain over time</td>
<td>Complete clinical information and good chronology</td>
</tr>
<tr>
<td>5. Patients surveys</td>
<td>Low degree of standardization in patient survey tools, often even within countries Cultural influences on concepts like satisfaction, expectations and experience hinders comparability across countries</td>
<td>Most reliable method of assessing system responsiveness and obtaining information about how patients perceive and experience the care provided Leads to improvements in designing trans-cultural assessment tools</td>
</tr>
<tr>
<td>Satisfaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
screening for each available year were calculated separately for programme and survey data. Based on surveys compared to registries, the variation over time is remarkable and suggests that both sources of data should be utilized with caution. Furthermore, international health system comparisons should use the source factor to adjust differences in the indicators.

**Identifying nationally representative data**

Cross-national assessments should reflect country-wide data. This is especially true when using process indicators (e.g. measuring care for chronic diseases) where data are often derived from pilots or ad hoc registries and raises serious concerns about the representativeness of data. Unique patient identifiers could make patients much more traceable within routinely collected information and thereby increase the reliability of data collected.

To ensure data comparability across countries, the HCQI project recently adopted a system of classification of the quality of data. This system comprises three levels:

- A—corresponds to national administrative registries, with demonstrated non-selection bias;
- B—accounts for non-national administrative registries with demonstrated non-selection bias;
- C—applies to ad hoc registries (e.g. research and pilots) and any other source not classified elsewhere.

Such a system has the advantage of enabling data collection at different levels of quality and using all available data sources, while preserving the rigour of the analysis. For instance, only data within categories A and B can be utilized but C type data can be collected and efforts made to raise them to the two higher categories.

**Determining retrospective completeness of the time series**

Almost all international comparative efforts face problems in obtaining uninterrupted, reliable data over a given time period. This limits the validity of trend analysis and affects the ability to interpret related indicators together. The time lag between policy implementation (e.g. breast cancer screening for a target population) and expected outcomes (improvement in breast cancer survival rates) can hardly be accounted for in the absence of time series. Prospective time series rely on regularly updated, sustainable data sources; retrospective completeness could be hindered by problems with (for example) the availability of data that need to be considered during international comparisons.
International health system comparisons: From measurement challenge to management tool

Comparative projects of health systems similar to those developed and implemented by the OECD have great potential in driving health policy. There can be numerous methodological barriers but the process of identifying and overcoming these pitfalls can lead to valid, reliable conclusions that enable effective health decision-making for overall system improvement.

**Turning international health system comparisons into health system performance management**

International comparisons of health systems can be a valuable tool for governments to revise their policies, review accountability agreements, and reassess resource allocation procedures. However, to strengthen health systems it is necessary to use these comparisons for performance management purposes and, as a first step, to integrate performance data needs into the policy-making process. An example from the Ontario Ministry of Health and Long-Term Care illustrates the systematic use of performance information and its flow through the decision-making cycle (Figure 4). The diagram shows that comparative data can be used at different stages of the health ministry’s business cycle, which (as a continuous improvement process) facilitates the use of strategic performance information for performance improvement purposes.

**Figure 4. Conceptualizing the range of potential impacts of health system performance comparisons on the policy-making process**

Source: Veillard et al [18]
Similar examples can be found in the United States Veterans Health Administration where performance indicators were used to monitor the effects of health system reforms while driving accountability agreements at sub-system and individual levels [23]. Other successful case studies range from health care organizations [48] to private industry [15]. In order to guide health policy-makers in the delivery of better results, it is critical to turn strategy-based performance information into performance management systems.

**Translating performance information for policy-makers**

Another crucial aspect of performance management is translating performance information to make it simple and clear to policy-makers [49]. For instance, the Ontario Ministry of Health and Long-Term Care represented health system performance measures from two different perspectives: variation in performance over time and against selected benchmarks (or comparators), respectively. These approaches are interesting examples of how to present performance information to health policy-makers in relevant ways. For instance, Figure 5 indicates to Ontarian decision-makers whether performance is improving; if it is favourable compared to pre-defined benchmarks (standards, international comparators, provincial comparators); and the policy actions required for different levels of performance. This approach suffered from standardization difficulties but with comparable performance data can be a promising practice for governments wishing to benchmark their health system performance in a concrete fashion.

**Figure 5. Translating benchmarking information to policy-makers. Example from the Ministry of Health and Long-Term Care, Ontario, Canada [50]**

[Diagram showing performance indicators and their translating to policy-makers' decision-making process]
Funnel plots are another tool for benchmarking performance management and are used increasingly by countries such as the United Kingdom and Canada [51]. Figure 6 shows a set of funnel plots that represent the performance indicator (in this case, the rate of hospitalizations for ambulatory care sensitive conditions) with deviations from the average. A trend component is incorporated, by using an arrow to indicate whether performance has improved or declined; the length of the arrow shows the relative magnitude of change over time. The calculation of funnel plots is associated with some statistical problems but they can provide policy-makers with a visual representation of their country’s relative performance against comparators that is easy to interpret and helps to identify areas for improvement [51].

Figure 6. Funnel plots for ambulatory care sensitive conditions for different Canadian provinces, 2006 data [50]

Benchmarking Health System Performance

Despite the methodological difficulties of comparative efforts, the diversity of benchmarking initiatives shows that national and regional health authorities are gaining increasingly from comparing their performance and learning policy lessons from better performers. The selection of benchmarks is becoming more pragmatic and increasingly driven by the specific strategies of health systems and by performance expectations. Performance measurement thus becomes the basis for policy discussions concerning how to improve health system performance and specifically about sharing how others have achieved higher performance in a particular context. For instance, a number of European countries have invested in efforts to benchmark their performance against countries such as Australia, Canada, New Zealand, and the United States through the work of the Commonwealth Fund (Box 4).
Box 4. Benchmarking for better health system performance: example of the Commonwealth Fund in the United States

The Commonwealth Fund, a private organization in the United States, established the Commission on a High Performance Health System in 2005. This group of experts was assembled to analyse best practices from several health systems. Their benchmarking shows that Denmark performs better than any other country in Europe on measures of patient satisfaction and primary care; Germany is a leader in national hospital quality benchmarking; and the Netherlands and the United Kingdom lead on transparency in reporting quality data [52].

Within the United States, the Commission also benchmarked states against each other across five key dimensions of health system performance—access, quality, avoidable hospital use and costs, equity, and healthy lives [53]. Cumulative and dimension-specific ranks were published along with an analysis of the policy implications. The results are publicly available and are intended to assist states to identify opportunities to better meet the population's health needs and learn from high-performing states [54].

From this perspective, a well-designed benchmarking system has the potential to guide policy development and can be used both prospectively and retrospectively [2]. It can support better understanding of past performance and the rationale behind certain performance patterns (retrospective use) and also help to revise strategies for improving future performance (prospective use).

Such strategy-based performance benchmarking systems have certain characteristics:

• Strategic focus: link between health system strategies and international benchmarking efforts ensures that policy lessons will be designed for those who can act upon the findings (the policy-makers).
• Adaptability and flexibility: benchmarking efforts can undertake both large (full health systems comparisons) and narrower scope studies, using tools that can be administered in a time frame that matches the agenda of policy-makers (e.g. using patient survey comparisons such as that of the Commonwealth Fund).
• Data standardization: efforts are made to standardize data and facilitate credible comparisons.
• Policy focus rather than research focus: benchmarking systems are not driven by experts or researchers but by policy-makers supported by experts and researchers.
International health system comparisons: From measurement challenge to management tool

• Efforts to translate performance information and policy lessons for decision-makers: new tools (e.g. funnel plots) are used increasingly to represent performance information in rigorous yet explicit ways, conveying data in a meaningful manner while reducing the need to rank health systems in league tables.

• Sensitivity to political and contextual issues: interpretation of indicator data should not lose sight of the policy context within which they are measured; of the players involved in formulating and implementing policy; of the time lag needed to assess the impact of different policies; and of aspects of health care that remain unmeasured by available data.

Conclusion

This chapter reviewed the reasons for an increased governmental interest in international health system performance comparisons—they offer greater accountability and transparency and support strategy review and development. However, mutual learning is a third function that is becoming more important with the increasing scientific robustness of knowledge created through health systems research. Projects such as the OECD HCQI project or the Commonwealth Fund’s cross-national benchmarking initiatives in the United States are two good examples of comparative efforts in this direction. The scope of experiences is growing and covers comparisons at different levels of the health system and from different perspectives. The methodological difficulties of such exercises can be classified and addressed over time but require investment from countries. Governments can achieve superior health system performance through the powerful policy instruments offered by linking performance measurement to performance management; translating performance information in ways that are meaningful for policy-makers; and investing in benchmarking and mutual learning. Finally, important requirements for fostering the value of international comparisons and their practical use for performance improvement are listed below:

• Recognize the value of information and make substantial investments in improving minimum data quality for developing and transition countries (e.g. through the Health Metrics Network) and data quality for developed countries (through projects such as the OECD HCQI project).

• Build upon knowledge of how to resolve methodological issues in health system performance comparisons in order to strengthen such comparisons.

• Encourage international organizations to actively provide support for data standardization efforts within their member states.

• Achieve a balance between process and outcome indicators in comparisons of health system performance in order to provide different but complementary insights into health care processes.
• Avoid inconsistencies, strategic misalignment, and (ultimately) health system sub-performance by selecting indicators that cascade across different (macro, meso, micro) levels of the health system through performance measurement and accountability mechanisms.
• Benchmark networks structured against common strategic objectives and performance patterns to build stronger analytical capacities are built within and between countries.
• Evaluate indicator data across countries with an adequate understanding of the regulatory and evaluative policies that underpin them.
• Develop and use graphic tools to convey performance information to policy-makers in a meaningful way.
• Undertake further research in health system performance management and share the results effectively among countries.
International health system comparisons: From measurement challenge to management tool

References


