This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL REPORT

International benchmarking of healthcare quality

A review of the literature

Ellen Nolte

Prepared for the Department of Health
within the PRP project "An 'On-call' Facility
for International Healthcare Comparisons"

RAND EUROPE

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

The research described in this report was prepared for the Department of Health within the PRP project "An 'On-call' Facility for International Healthcare Comparisons (grant no. 0510002).

Published 2010 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
Westbrook Centre, Milton Road, Cambridge CB4 1YG, United Kingdom
RAND URL: http://www.rand.org
RAND Europe URL: http://www.rand.org/randeurope
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

This report reviews recent developments in the field of international benchmarking of healthcare quality to inform the use of international comparisons for quality improvement in the NHS.

The report was prepared as part of the project "An 'On-call' Facility for International Healthcare Comparisons" funded by the Department of Health in England through its Policy Research Programme (grant no. 0510002).

The project comprises a programme of work on international healthcare comparisons that provides intelligence on new developments in other countries, involving a network of experts in a range of OECD countries to inform health (care) policy development in England. It is conducted by RAND Europe, in conjunction with the London School of Hygiene & Tropical Medicine.

RAND Europe is an independent not-for-profit policy research organisation that aims to improve policy and decision-making in the public interest, through rigorous research and analysis. RAND Europe's clients include European governments, institutions, NGOs and firms with a need for rigorous, independent, multidisciplinary analysis.

The London School of Hygiene & Tropical Medicine is Britain's national school of public health and a leading postgraduate institution worldwide for research and postgraduate education in global health.

This report has been peer-reviewed in accordance with RAND's quality assurance standards.

For more information about RAND Europe or this report, please contact:

Dr Ellen Nolte
Director, Health and Healthcare
RAND Europe
Westbrook Centre
Milton Road
Cambridge CB4 1YG
United Kingdom
Email: enolte@rand.org
Tel: +44 (0)1223 273853

Contents

Figures

Tables

Summary

There is growing interest in the systematic assessment and international benchmarking of quality of care provided in different healthcare systems, and major work is under way to support this process through the development and validation of quality indicators that can be used internationally.

Efforts under way involve systematic analysis of the suitability of routine datasets for comparing quality of care in different countries, including the development of standard definitions to improve the comparability of national data systems. There is scope to develop indicators which are specifically designed to compare care quality in different systems. The initiatives of greatest relevance to England are probably those seeking to benchmark quality of care and outcomes in European countries.

Recognising that cross-national data comparison remains a challenge, there is now a considerable body of data that allow for comparisons of healthcare quality in selected areas of care. The report includes a description of existing indicators that could be used to compare healthcare quality in different countries, along with a discussion of specific problems in making comparisons at this level of detail. This is illustrated with case studies of two measures widely used for international comparisons: avoidable mortality and cancer survival. These show both the potential power of cross-national comparisons and some of the difficulties in drawing valid interpretations from the data.

The report focuses on the three quality domains identified as important by the NHS Next Stage Review *High Quality Care for All*, namely effectiveness of care, patient safety and patient experience. It is however important to recognise that access is an important additional component of quality which may be a key determinant of differences in outcome between different countries. Thus comparing quality across countries is only a first step to then assess the causes underlying those differences and determining what actions may be appropriate to take to improve health outcomes.

International benchmarking of quality of care with the NHS has considerable potential to improve patient outcomes in England, but only if research is undertaken to understand the reasons for differences between countries and their possible relevance to England.

Acknowledgements

The project "An 'On-call' Facility for International Healthcare Comparisons" is funded by the Department of Health in England through its Policy Research Programme (grant no. 0510002).

The author gratefully acknowledges the valuable comments provided by Martin Roland and Tom Ling on an earlier draft of this report.

The views expressed in this report brief are those of the author alone and do not necessarily represent those of the Department of Health. The author is fully responsible for any errors.

Introduction

The NHS Next Stage Review *High Quality Care for All* set out an ambitious vision for making quality improvement the organising principle of the NHS.[1] As part of this vision, it has made a commitment to measure quality at the various levels of the system within a national quality framework. The framework includes indicators for quality improvement developed in partnership with the NHS for local clinicians to use to drive quality improvement in their area through measurement. There is also an international component to enable performance comparison at the system level with other advanced healthcare systems. The newly formed National Quality Board (NQB) has been tasked with overseeing the work to improve quality indicators and to examine how to use international comparisons for quality improvement.[1]

This report seeks to inform this process, focusing in particular on the development of internationally comparable measures on healthcare quality for potential use in the forthcoming NQB quality report to enable comparison at the system level. It does so through (i) providing a summary overview of conceptual and methodological considerations regarding the measurement of healthcare quality and the development of (international) quality indicators; (ii) reviewing the literature on existing international or national initiatives that employ international comparisons of metrics of healthcare quality; and (iii) assessing the suitability of existing indicators for international healthcare quality comparison, focusing on the three dimensions of healthcare quality defined by *High Quality Care for All*: effectiveness of care, patient safety and patient experience. The emphasis will be on national-level indicators that are considered to measure system-level progress. The overarching aim of this report is to inform on recent developments in the field of international benchmarking of healthcare quality; a detailed discussion of the advantages and disadvantages of existing approaches is beyond the scope this work. The review draws to a considerable extent on previous comprehensive reviews and related work undertaken by the authors that have been published elsewhere.[2-7] It is complemented by a review of the published and grey literature on national and international initiatives and performance assessment frameworks that use international benchmarking of healthcare, retrieved through an iterative search using bibliographic databases, common worldwide web search engines and specific government and agency websites.

Measuring healthcare quality: conceptual and methodological considerations

Improving the quality of healthcare can be achieved without actually measuring it, for example, through supporting the use of guidelines or peer review.[8] However, measurement has an important role in quality improvement, as a means to monitor effectiveness, protect patient safety, inform decision-making and ensure value for money, among many other purposes.[9] At the same time, identifying meaningful measures suitable to capture the quality of care in its different dimensions remains a challenging aspiration.[10]

There has been considerable work into the development and use of quality indicators,[a] which has been reviewed in detail elsewhere.[8-15] We here summarise the published evidence on three aspects that appear relevant in the context of indicator development for international comparison: the type of measures used; the desired attributes of quality indicators; and the use of routine data to assess quality, which we will address in turn.

Process or outcome measures?

While most approaches to monitor the quality of care tend to use a combination of different types of measures, aiming to assess structure, process and outcomes, there is an ongoing debate on the relative usefulness of process versus outcome measures to evaluate healthcare quality.[16]

Process measures offer an important tool for assessing the current quality of care being delivered by a system or in a country; they are useful for evaluating whether interventions have led to improved quality of care.[15] Thus process indicators:[11 15 16]

- tend to be more sensitive to changes in the quality of care and provide a clear direction to identify what needs to be done differently to achieve optimal care delivery;

[a] Some authors refer to 'performance' indicators to denote a similar concept although 'quality' and 'performance' are not necessarily identical, depending on the definitions being applied. The most common definition of 'quality' used in the literature on quality indicators is that proposed by the US Institute of Medicine (*Degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge*) (1992) whereas 'performance' is understood as a broader, multidimensional concept that also includes dimensions of equity and efficiency (e.g. Girard & Minvielle 2002). However, frequently the notions of quality and performance are used interchangeably and consequently, while recognising the conceptual differences, in this report, we will use the terms interchangeably also.

- are easily measured without major bias or error (i.e. they do generally not require risk adjustment) and are easy to interpret;
- enable detection of deficits in care more rapidly compared to outcome measures as care processes occur more frequently (i.e. a higher number of cases allows for more robust assessment); this also means that the sample size required for detecting quality deficits will be much smaller compared to outcome measures;
- capture aspects of care that are important to patients (in addition to health outcomes), such as speed of access.

At the same time, however, process indicators are very specific as they tend to focus on a given intervention or condition; as a consequence, a range of process measures may be needed in order to assess the quality of care for a particular group of patients. In addition, indicators may quickly become outdated as organisations and technology advance, or they may indeed stifle innovation.[11] Finally, process indicators can easily be manipulated, so potentially undermining quality improvement efforts.

Outcome measures focus "the attention of policy makers on whether systems are achieving the desired goals"[17] and towards the patient (rather than the service). An emphasis on outcomes may promote innovation through for example encouraging providers to experiment with new models of care to improve patient outcomes as well as supporting the adoption of long-term health promoting strategies.[11] Also, the ability to measure outcomes is necessary to ensure that process changes actually improve attainment of desired health system goals.[17]

However, there are numerous challenges towards the interpretation of outcome measures. One relates to the observation that the outcomes of interest are often (much) delayed and it is thus difficult to establish a clear link to a given intervention.[18] One example is the use of survival data to monitor the outcomes of cancer care. Typically, these data are only measured several years later, so reflecting previous clinical practice, which will be difficult to influence in terms of quality improvement (see also below). Also, there may be challenges to attribute observed change as it is not always clear why outcomes are poor and it may require collection of process measures to identify steps that should be taken to improve outcomes.[15] Furthermore, the collection of data on outcomes may take considerable time and typically requires sizeable numbers in order to detect statistically significant effects even when there are obvious shortcomings with the processes of care.[11]

In summary, both process and outcome indicators have merits and risks which need to be taken into account when developing measures to assess the quality of care. Mant (2001) noted that the relevance of outcome measures is likely to increase with the broadening of the perspective, i.e. towards macro-level assessments of quality because such measures tend to reflect the inter-play of a range of factors, some of which directly related to healthcare[16] and these factors are more easily addressed at the national or system level. Conversely, at the organisational or individual team level, process measures will become more useful as broader outcome measures are less easily influenced at this level.

Desirable attributes of (international) quality indicators

The term 'quality indicator' has been defined in different ways,[8 14 19] however, the common notion is that an observed change in a given indicator reflects something about the underlying healthcare delivery and the quality of care.[11] Therefore, if quality measurement is to guide quality improvement, indicators should meet certain criteria to allow for appropriate conclusions about cause and effect to be drawn and/or cause of action to be taken. Frequently, however, quality measures seem to be selected on the basis of what is available and practical ('measurable') rather than what is meaningful,[20] and the extent to which these reflect the quality of healthcare, or indeed their implications for policy, remain unclear.

In response, analysts have presented lists of desirable attributes of quality indicators, with *validity* (the extent to which the measure captures the concept it is meant to measure), *reliability* (the extent to which measurement with the given indicator is reproducible) and *sensitivity to change* considered among the key criteria.[11] Depending on the context and purpose of measurement, the range of indicator attributes may be broadened, however. For example, Pringle et al. (2002) proposed a list of 12 attributes of quality indicators to guide indicator selection, arguing that these should, in addition to being valid and reliable, also be communicable, effective, objective, available, contextual, attributable, interpretable, comparable, remediable and repeatable,[10] with others adding adaptability,[21] feasibility,[22] acceptability,[8] policy relevance[22 23] and actionability[24] as further criteria for quality indicators. Table 1 presents an overview of indicator attributes and common definitions used.

The applicability and relevance of the criteria listed in Table 1 will vary with the purpose and context of measurement. For example, international quality measurement initiatives have, to a great extent, to rely on existing data sets to enable comparison. Thus feasibility is likely to form an important criterion for indicator selection, as for example in the Commonwealth Fund International Working Group on Quality Indicators initiative, which selected performance indicators based on five criteria: (1) *feasibility* (indicators are already being collected by one or more countries); (2) *scientific soundness* (indicators are reliable and valid); (3) *interpretability* (indicators allow a clear conclusion for policy-makers); (4) *actionability* (measures can be directly affected by the healthcare system); and (5) *importance* (indicator reflects important health conditions in terms of burden of disease, cost of care or priorities of policy-makers) (see also below).[24]

Table 1 Quality indicator attributes and descriptions

Attribute	Description
Valid	The extent to which a measure accurately represents the concept/phenomenon being evaluated
Communicable	Relevance of measure can be easily explained and understood by target audience
Effective	Indicator measures what it aims to measure; free of perverse incentives
Reliable	The extent to which a measurement with an indicator is reproducible
Objective	Data are independent of subjective judgement
Available/feasible	Data are collected for routine (clinical/organisational) reasons and are available quickly and with minimum extra effort or cost
Contextual	Measure is context-free, or important context effects should be adjusted for
Attributable	Measure reflects the quality of care delivered by individuals, teams and organisations
Interpretable	Measure allows for ready interpretation of core underlying factors
Comparable	Measure allows for reliable comparison with external benchmarks or to other datasets collected in similar circumstances
Remediable/actionable	Measure points to actionable areas for improvement that are likely to impact positively on the measure in question
Repeatable	Measure is sensitive to improvements over time
Adaptable	Measure is appropriate for use in a variety of contexts and settings
Acceptable	The extent to which the process of measurement (and reasons for it) is accepted by those affected
Policy-relevant	Indicator reflects important health conditions in terms of burden of disease, cost of care or public interest

Source: adapted from [8 10 11 14 22 23]

A key challenge to indicator selection is that indicators should be evidence-based; however, the available evidence is not necessarily strong in many areas except the most studied, as for example in the case of angina. As a consequence, performance/quality indicators will have to rely to a larger extent on expert opinion.[8 11] This may cause problems when different professional groups with different perspectives on what constitutes 'high quality care' are being consulted (and so can the actual consultation process).[25] Further challenges may arise in relation to the transferability of quality indicators between countries, because of differences in professional opinion and in the interpretation of evidence; also the evidence-base used might vary, for example building on evidence that is available in the native language of one country only.[26-28]

Use of routine data to assess quality of care

Limited data availability and lack of uniformity of data across different settings pose substantial challenges to most initiatives seeking to assess the quality of healthcare systems.

The use of routine data for quality measurement has many advantages as data are readily available in many healthcare settings.[18] Also, data can be used retrospectively and so enable compilation of data sets extending over longer time periods. Routine data are typically low-cost as the data are being collected already for purposes other than quality assessment. Data are usually observational, capturing processes of care (e.g. treatment given, length of hospital stay, vaccination rates, screening rates etc.) as well as health outcomes such as mortality (e.g. perioperative deaths) and intermediate outcomes such as physiological measures (e.g. blood pressure). In addition, routine data provide a potentially rich source of information on a large number of patients with different health problems across diverse settings and/or geographical areas.

At the same time, however, there are considerable threats to the validity and reliability of routine data, which Powell et al. (2003) have summarised as follows:[18]

- *Data quality*: data may be incomplete or inaccurate.
- *Level of detail*: routinely collected data such as administrative data often lack the level of clinical detail required to assess quality of care (i.e. may be insensitive to capturing instances of substandard care).
- *Insensitive/inappropriate definition of outcomes*: e.g. '30-day in-hospital deaths' are generally defined as deaths in the admission in which a given intervention was performed while deaths that occur elsewhere (outside the hospital where the intervention was performed, e.g. at home or another hospital after transfer) are excluded even though they may still be related to the care provided in the admitting hospital. This measure can be particularly problematic in the light of shorter length of hospital stay and an increased tendency to transfer patients between hospitals for specialist care; as a consequence, a greater proportion of deaths within 30 days of admission will be missed if only 'in-hospital deaths' are being considered.
- *Measures are too narrow* to reflect the care provided. Thus Giuffrida et al. (1999) demonstrated how using admission rates for certain conditions (asthma, diabetes, epilepsy) as an indicator of the quality of primary care (based on evidence that high quality primary care tends to be associated with lower admission rates for these conditions) tends to be misleading.[29] Admission rates can be interpreted as a (crude) indicator for population health/health outcomes; however, admission rates are influenced by a range of factors, including population characteristics such as socioeconomic profile and morbidity ('case-mix') and secondary care characteristics which are usually outside the control of primary care practices. While adjusting rates to take account of these factors may improve their applicability as a measure of the quality of primary care, it only addresses one of the many dimensions of quality care.
- *Changes in reporting practices over time*: may suggest improvement/deterioration in a given indicator although any change is most likely related to changes in data recording.
- *Differences in data collection methods* between providers/regions/systems may suggest variation on a certain indicator although variation is more likely caused by differences in the data collection.

The two last points, and data comparability more generally, are of particular relevance for international comparisons. The Organisation for Economic Co-operation and Development (OECD) has undertaken pioneering work in assembling an international database of inputs, processes and outcomes of health systems.[30] In doing so, it has identified many weaknesses in the existing data. For example, figures for numbers of health professionals in some countries were based on head counts, taken from professional registers, while in others they were limited to those in employment (and in some cases, only those working in the state sector).[7]

Accurate collection of indicator data relies on the existence of reliable and well-established health information systems. However, most existing health information systems were originally devised for internal mechanisms of financial control, and their adaptation for purposes of quality assessment may not be easy. Problems with minimum data sets, inaccuracies in interpretation of aggregated data, failure to integrate population- and patient-level data and lack of linkage between diagnostic data and outcomes of care are some of the main drawbacks reported in existing health information systems.[31]

Overview of (inter)national comparisons of healthcare quality

The comparative assessment of the performance of healthcare systems within and between countries has received growing interest, in particular over the past 10 years or so. Early examples include the work of the OECD on international benchmarking of health systems, through a series of international studies published from the mid-1980s[32] and with a major focus on inputs into healthcare such as healthcare expenditure and human resources, and the World Health Report 2000 and its ranking of the world's health systems.[33] The latter in particular has stimulated wide-ranging debate about approaches to assessing health system performance both nationally and internationally.[34]

Following these debates, there has been a shift away from an emphasis on healthcare inputs towards outcome orientation. Also, recognising that different models of healthcare organisation and provision produce varying results, both overall and with different balances across health sectors, there has been a shift away from using composite performance indices, such as those used in the 2000 World Health Report, towards the development of methods and techniques that are suitable to capture these differences in a systematic and comparable way. More recent approaches and frameworks to assess healthcare quality and/or performance at the macro (i.e. national and/or international) level therefore use a range of indicators so as to capture the different aspects of healthcare systems.[35] Examples include the Commonwealth Fund International Working Group on Quality Indicators initiative and the subsequent OECD Health Care Quality Indicator Project (see below).[24 36]

At the national level, several countries have been developing conceptual frameworks for monitoring and assessing the performance of their health systems with improving the quality of care at the core of many such initiatives. The earlier initiatives, such as the 1998 Department of Health's NHS Performance Assessment Framework,[37] the 2000 Canadian Health Information Roadmap Initiative Indicators Framework,[38] the 2001 Australian National Health Performance Framework,[39] (ongoing) work by the US Agency for Healthcare Research and Quality (annual National Health Care Quality Report published from 2003 onwards),[40] along with the early work by OECD and the World Health Organization mentioned above have been described in detail elsewhere.[35]

We here explore more recent macro- or system-level performance or quality assessment initiatives that are either international by inception or draw, to a considerable extent, on cross-national comparisons. This overview is of necessity highly selective: several countries

have now developed national quality assessment frameworks, such as Denmark,[41] the Netherlands,[42 43] Sweden[44 45] and Taiwan,[46] while others have renewed or advanced existing frameworks, such as Canada's Health Indicator Framework (2003),[47 48] the Quality and Outcomes Framework (QOF) in England,[49] and recent work in Australia to strengthen the accountability of the system through benchmarking.[50 51] However, few national initiatives explicitly include international comparison of quality indicators; indeed, among those listed, only the Dutch approach systematically considers international data for benchmarking selected areas.

In addition to these (inter)national level activities that are aimed at assessing the quality of care at the system level, there are several initiatives or projects that, while not specifically addressing overall health system quality and performance, have undertaken developmental indicator work and/or provide data that can inform and/or are being used in international healthcare comparisons. Several such initiatives are at the research level, focusing on disease-specific areas such as the EUROCARE study (cancer)[52 53] or the EUCID project (diabetes);[54] the development of European level (public) health indicators such as the ECHIM project[55 56] and the ISARE project;[57 58] the development of quality indicators for primary care in selected European countries;[59 60] and the collection of cross-national data on patient experience such as the Commonwealth Fund international health policy surveys,[61-63] focusing on experience of patients with chronic health problems, with the most recent survey covering eight countries,[64] and the World Health Survey.[65 66] In addition, databases such as the World Health Organization Regional Office for Europe Health for All database[67] and the OECD Health database[68] provide collections of a wide range of data on health and healthcare in a large number of European and/or OECD countries. While it is beyond the scope of this report to review these projects and initiatives, we will reflect on selected sources in relation to their suitability for international benchmarking of healthcare quality at the system level.

We reflect on two national healthcare quality assessment initiatives: the Dutch Health Care Performance Report,[42 43] commissioned by the Dutch government, and the National Scorecard on US Health System Performance (National Scorecard),[69-71] an initiative by the Commonwealth Fund Commission on a High Performance Health System; as well as two international initiatives: the World Health Organisation's Health Systems Performance Assessment Framework (WHO HSPA)[33 72 73] and the OECD Health Care Quality Indicators Project (HCQI).[22 74 75] Initiatives are examined according to origins, aims and scope, as well as criteria for indicator selection and operationalisation of indicators, shown in Table 2 and Table 3 respectively.

At the outset it is important to note that the WHO HSPA, and its application to national health systems as presented in the World Health Report 2000,[33] does not represent a quality assessment framework per se; its focus is on health system performance, which is interpreted as attainment of three system goals: health improvement, responsiveness to expectations and fairness in financial contribution. Goal attainment is assessed through five indicators, namely the overall level of population health and its distribution in the population; the overall level of responsiveness and its distribution within the population; and the distribution of the health system's financial burden within the population. Each is set against the resources available and brought together as a composite index as a measure of 'overall attainment' (performance).[72] Quality is understood as a subset of overall goal

attainment and assessed as average levels of population health and of responsiveness.[73] It is further worth noting that the WHO HSPA as a periodic assessment of health system performance of WHO Members States has been discontinued; however, further developmental work on for example health system responsiveness is ongoing.[65]

The OECD Health Care Quality Indicator Project (HCQI) builds, to considerable extent, on earlier work by the Commonwealth Fund sponsored International Working Group on Quality Indicators (CWF QI) mentioned earlier, and the Nordic Indicator Group Project set up by the Nordic Council of Ministers (Table 2). HCQI's origins and history have been described in detail elsewhere.[76] However, for the purposes of this report it is important to highlight that its work has been evolving rapidly, and continues to do so, both in scope and number of participating countries. The project involves a substantial methodological component in terms of indicator development and evaluation, along with assessments of the feasibility to collect internationally comparable data that can be released publicly. As a consequence of its emphasis on methodological work, the public release of international data has so far been limited to a small set of indicators covering selected areas in healthcare.[77]

The 2008/09 HCQI data collection now includes 40 quality indicators, covering prevention and primary care, selected areas of curative care, patient safety and mental health (Table 3). The number of participating countries has grown from 23 in 2003 to 32 in 2007.[36]

In contrast to HCQI, with its emphasis on a comparatively small set of quality indicators in a few select areas, largely driven by the availability of comparable data suitable for international comparison, the Dutch Health Care Performance Report (DHCPR) and the National Scorecard on US Health System Performance (National Scorecard) use international comparative data as a means to benchmark national data in addition to (or in the absence of) a given national benchmark (however defined) (Table 2). However, similar to HCQI, indicator selection, whether national or international, is driven by feasibility criteria, i.e. whether data are available and readily accessible. It is noteworthy, though, that both DHCPR and the National Scorecard include access and patient experience as a key domain of healthcare quality, in both cases benchmarking against international data (see below), while the current focus of HCQI is on effectiveness and patient safety, although patient-centeredness/responsiveness is recognised as a key attribute of healthcare quality (Table 3), with related development work under way.[78]

Table 2 Selected healthcare quality/performance assessment initiatives that include cross-national comparisons: origins, aims and scope

Initiative	Stated aim(s)	Origins and periodicity	Setting/countries involved or presented	Target audience
WHO Health Systems Performance Assessment Framework (WHO HSPA)[33 72 73 79]	*Focus*: health system performance including health attainment, responsiveness and equity *Aims*: • To develop a framework for describing, analysing and improving health system performance • To develop effective and affordable tools that can be used nationally to provide timely and relevant information on system performance • To undertake periodic assessments of health systems performance in the WHO Member States	1998; development of conceptual framework and its application to national health systems presented in World Health Report 2000; discontinued as a periodic assessment of health system performance of WHO Members States; ongoing development work on e.g. health system responsiveness[65]	191 WHO Member States (2000)	National decision-makers
OECD Health Care Quality Indicators Project (HCQI)[22 36 74 75]	*Focus*: "technical quality of health care" with effectiveness, safety and responsiveness/patient centeredness as core quality dimensions[74] *Aims*: • To "develop a set of indicators that reflect a robust picture of health care quality that can be reliably reported across countries using comparable data"[74] • To "develop a set of indicators to raise questions about health care quality across countries for key conditions and treatments"[75] • *Secondary goal*: "to support efforts aimed at coordination between major international organisations seeking to track health care quality indicators" and so to reduce the data collection burden on participating countries and to improve data comparability internationally[75]	Builds on work by Commonwealth Fund[24] and the Nordic Indicator Group Project[80] Report on first round of data collection covering 2003-2005 published in 2006; updated in 2007; further indicator development work ongoing	2006: 23 OECD countries 2007: 32 countries (incl. two non-OECD EU countries)	Not explicit
Dutch Health Care Performance Report (DHCPR)[42 43]	*Focus*: Performance of the Dutch healthcare system in terms of three system goals: quality, access and costs, from a macro perspective *Aims*: • To "provide a picture of the performance of Dutch health care at the national level, based on a limited set of indicators and using existing data"[43] • To enable "policy appraisal of health care performance" through time trends, international comparisons and comparison with policy norms and objectives • To "monitor the quality, accessibility and efficiency of Dutch health care" so as to "contribute to strategic policy decisions regarding health care"[42]	Production of the report commissioned by the Dutch Ministry of Health, Welfare and Sport every 2 years from 2004 (English edition: 2006) and once every 4 years from 2010	The Netherlands Varying range of OECD countries used to benchmark certain areas (for example where there are no national (policy) standards)	Not explicit
National Scorecard on US Health System Performance (US National Scorecard)[69 71]	*Focus*: Performance of the US healthcare system; measuring US health system performance against specific benchmarks *Aims*: • To provide benchmarks for assessing health system performance • To establish a mechanism for monitoring change over time • To enable estimating the effects of proposed policies to improve performance	Initiated in 2005 by the Commonwealth Fund through establishment of a Commission on a High Performance Health System Publication of two reports since (2006, 2008)	USA Varying range of OECD countries used to benchmark certain areas	Not explicit

Table 3 Selected healthcare quality/performance assessment initiatives that include cross-national comparisons: indicator frameworks

Initiative	Definitions	Criteria for indicator selection	Indicator development/application
WHO HSPA	• *Health system*: all actors, institutions and resources that undertake health actions* are considered part of the health system • *Performance*: extent to which a health system achieves the desired outcomes given available resources • *Quality*: not explicit; quality interpreted as a subset of overall goal attainment (average levels of health and responsiveness) • *(Quality) indicator*: not defined * primary intent of a health action is to improve health	Oriented at primary goals of a health system: o improving population health o improving responsiveness o fairness in financial contribution and four key functions contributing to achieve goals: o financing o service provision o resource generation o stewardship	Five main indicator framework reflecting the three major health system goals with their average levels and distribution: 1. Health improvement (two indicators: level and distribution) 2. Responsiveness to expectations (two indicators: level and distribution) • respect for persons (dignity, individual autonomy, confidentiality) • client orientation (prompt attention to health needs; basic amenities; choice) 3. Fairness in financial contribution (one indicator: distribution) Goal attainment ('performance') is measured for each indicator and as a composite index from the weighted sum of the components ('overall indicator of attainment')
HCQI	• *Health (care) system*: not explicit • *Performance*: not explicit; multidimensional concept that includes quality as well as access and costs (not addressed by HCQI) • *Quality*: based on IOM definition "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge"; effectiveness, safety and responsiveness/patient centeredness form core attributes of quality[74] • *Quality indicator*: measure "of health outcome or health improvement attributable to medical care"[22]	Indicator selection criteria[36 75] • Importance: o Impact on health (*the measure addresses areas in which there is a clear gap between the actual and potential levels of health*) o Policy importance (*economic importance of the condition covered by the indicator*) o Susceptibility to be influenced by healthcare system (*degree to which the healthcare system can meaningfully address the health issue*) • Scientific soundness: o face validity o content validity o reliability Additional criterion: feasibility (i.e. data availability)[75]	Indicator development carried out in several stages and evolving: *Stage 1*: testing and evaluation for suitability for international comparison of initial set of 21 indicators which meet the indicator selection criteria in prevention and curative sector[75] *Stage 2*: development of additional indicators in five priority areas (cardiac care; diabetes care; primary care and prevention; mental health; patient safety) resulting in a list of 86 indicators considered valid and reliable; subset of 23 found to be actually available; subset of 5 included in the 2006 list of HCQI indicators (N=26; 7 of which deemed 'not fit' for international comparison)[36] *Ongoing*: further indicator development with particular focus on mental health and patient safety[81 82] 2008-09 HCQI data collection covers 40 indicators: • regularly collected indicators (12) o 8 outcome indicators (cancer, cardiovascular disease, asthma) and 2 process measures • health promotion, prevention and primary care indicators (9) • mental health (2) • patient safety (7) • set of 10 additional indicators addressing selected outcomes (vaccine-preventable disease; smoking prevalence) and processes measures (cancer screening; vaccination rates)

Table 3 Selected healthcare quality/performance assessment initiatives that include cross-national comparisons: indicator frameworks *(continued)*

Initiative	Definitions	Criteria for indicator selection	Indicator development/application
DHCPR	• *Healthcare*: "[A]ctivities aimed at alleviating, reducing, compensating and/or preventing deficiencies in the health status or autonomy of individuals"[43] • *Performance*: not explicit; multidimensional concept that includes three dimensions (quality, access, costs) along the continuum of care (prevention, cure, long-term care, palliative care) • *Quality*: based on the 2001 definition of the Institute of Medicine ("doing the right thing, at the right time, in the right way, by the right person') and the Dutch Quality of Care Institutions Act, which refers to 'responsible care', defined as "care of a high standard that is provided in an effective, efficient and patient-centred way and that meets the patient's actual needs"[42]* • *(Quality) indicator*: "a measurable aspect of care that gives an indication of a specific performance aspect"	Oriented at the three system objectives/goals: o quality o access o costs. Goals are subdivided into indicator *domains* based on the (international) literature; each indictor domain is further divided into indicators on the basis of having "a signalling function for the domain concerned". Indicator selection is based on: • the intrinsic relationship between the indicator and a specific care aspect to be investigated • data availability	Use of initially (2004) 15 indicator domains (2008: 12) to assess achievement of the three system objectives (quality, access, costs), further subdivided into a total of 125 (2008: 110) indicators *Domains and indicators of 2008 report* Quality domains • effectiveness (51 indicators) o prevention (8) o curative care (8) o long-term care (10) o mental health (5) • patient safety (10) • innovation in healthcare (10) Access domains (43) • financial access to care (6) • geographical access to care (5) • timeliness of acute care (6) • waiting for regular care (6) • access according to needs (5) • staff availability (10) • freedom of choice (5) Costs of care (16) • trends in health expenditure (8) • financial position of healthcare organisations and health insurers (5) • labour productivity in healthcare (3) Connecting themes • public & patient views on and experience of the healthcare system • efficiency of the healthcare system • market forces and the effects of the 2006 healthcare system reforms

Table 3 Selected healthcare quality/performance assessment initiatives that include cross-national comparisons: indicator frameworks *(continued)*

Initiative	Definitions	Criteria for indicator selection	Indicator development/application

US National Scorecard	• *Health (care) system*: not explicit
	• *Performance*: multidimensional concept; defines four core goals of a high performance health system "1) high quality, safe care; 2) access to care for all people; 3) efficient, high value care; and 4) system capacity to improve"[70]
	• *Quality*: a "broad measure covering the extent to which the care delivered is effective and well-coordinated, safe, timely, and patient-centered"[69]
	• *(Quality) indicator*: not defined
		Indicator selection criteria:[83]
		• policy importance: areas where improvement (of the healthcare system) can make a significant difference
		• feasibility: data is available and readily accessible from national or international databases
		• potential for time-trend analyses
		• enables comparisons of US average performance levels to benchmarks drawn from national and international experiences
		
		Indicator development/selection informed by indicators developed elsewhere (e.g. Agency for Healthcare Research and Quality, National Committee for Quality Assurance, Joint Commission on the Accreditation of HealthCare Organizations
			Indicators are organised into five domains/dimensions of performance: health outcomes, quality, access, equity and efficiency, using a total of 37 indicators:
			• health outcomes (5 indicators)
			• quality (19)
			  o the right care (5)
			  o coordinated care (5)
			  o safe care (4)
			  o patient-centred, timely care (5)
			• access (5)
			  o universal participation (2)
			  o affordable care (3)
			• efficiency (8)
			  o overuse/inappropriate care (1)
			  o access to timely care (2)
			  o variation in quality and costs (3)
			  o insurance administrative costs (1)
			  o information systems (1)
			
			Equity is measured across selected indicators of above domains (e.g. health outcomes and insurance coverage by race/ethnicity).
			
			Each indicator is scored relative to benchmarks of higher performance, with a maximum score of 100, based on the 'best' benchmark rates internationally ('top' countries) or the top 10 percent of US states/providers[83]

Suitability of existing quality indicators for international comparison

The HCQI indicator development process has been described in detail;[22 36 75 76] we here highlight a few features of that process that appear to be of particular relevance for international comparisons. For example, as shown in Table 3, the quality indicator selection process is very much guided by desired indicator attributes described earlier, including validity, reliability, sensitivity to change, as well as policy importance. Yet, the international dimension of the work means that methodological rigour will at times have to be balanced against 'inclusiveness'. Thus there will be cases in which a given indicator is not nationally representative but reflects a selected region, or locality in the region only. There is a persuasive argument to not exclude locality indicators given the overarching aim of the project to enable comparison of international quality of healthcare.[74]

One other challenge relates to reconciling the desire to address a wide range of disease areas and population groups, and the many possible interventions at different stages, against the level of detail, or comprehensiveness, with which a given disease or population group can be represented in relation to a set of interventions. The HCQI Ad Hoc Group on Health addressed this issue by recommending that the number of indicators should not exceed 50 in the first instance.[74] The reasoning behind choosing 50 indicators as a target remains unclear. At the same time, there appears to be a general preference among member countries for using fewer, more comparable indicators rather than a broader, less comparable set of indicators, suggesting that indicators are likely to be defined, at least initially, as much by the availability of data as by the priority which was accorded to the indicators.[74] There is an expectation that additional indicators will be included as data availability, quality and comparability improves. However, as indicated in Table 4, comparability of data remains a key challenge.

A further issue concerns determining the minimum number of countries for which data of appropriate quality on a given indicator that has met the indicator selection criteria should be available to allow for meaningful comparison. The OECD HCQI has stipulated a threshold of 10 countries that are able to provide data on a given indicator for inclusion in the HCQI indicator list.[36] As a consequence, while several indicators may have met the inclusion criteria on the basis of being scientific sound and addressing an important problem, if fewer than 10 countries can provide relevant data from well-defined and stable databases according to agreed definitions, the indicator will not be included.[84]

This last point highlights one of the key challenges to advancing further international benchmarking activities. As noted earlier, one important criterion for indicator selection is feasibility, i.e. data to be drawn on should be collected for routine reasons and be available quickly and with minimum extra effort or cost (Table 1). Yet, many indicators, in particular those that reflect patient experience, are often not available routinely but are typically collected through surveys. While several countries have instituted regular relevant instruments, such as the NHS patient surveys in England,[85] data are not necessarily comparable with similar surveys undertaken elsewhere, in particular when the data collection instrument cannot ensure cross-cultural equivalence.[78] Where cross-national comparable instruments have been employed, these frequently tend to cover a smaller range of countries only, often building on small samples of uncertain representative power in participating countries, and/or surveys are not undertaken regularly, or even only once, so data tend to become outdated. For example, the World Health Survey, which provides a rich source for assessing the patient experience in 70 countries,[65] was implemented in 2002-03, so reflecting the responsiveness of health systems at that time only. A related issue is the variability of data sources, both within and between countries, as noted earlier.[84]

Tables 4-6 provide an overview of quality indicators that are being used and/or suggested for use in international comparison, with Table 4 displaying indicators of effectiveness of care (broadly categorised according to disease area), Table 5 listing indicators of patient safety and Table 6 listing indicators of patient experience. Where appropriate and relevant, the tables include brief comments on the suitability of the corresponding indicator for international comparison. It is beyond the scope of this report to review each indicator in detail; indeed, a great many indicators are derived from the OECD HCQI, which has provided detailed documentation on methodological concerns around comparability.[36 75 76 81 82] However, we explore some of the issues raised by means of two 'case studies' that examine the methodological challenges to the use of (1) the concept of 'avoidable' mortality and (2) of cancer survival as a means to assess the quality of healthcare international comparison.

Table 4 International quality indicators: effectiveness (by disease area)

Indicator	P§	O§	Used by	Data source	Indicator of	Comment
Vaccine-preventable disease						
Vaccination rate measles, age 2	✓		HCQI	HCQI participating countries	Not specified	Judged suitable for international comparison[75]
Vaccination rate pertussis, age 2	✓		HCQI	HCQI participating countries	Not specified	As above
Vaccination rate hepatitis B, age 2	✓		HCQI	HCQI participating countries	Not specified	As above
Vaccination rates of DTP-3 (diphtheria, tetanus, whole cell pertussis)	✓		DHCPR 2008[42]	WHO Health for All database	Effectiveness of prevention	Benchmarked against 27 EU Member States
Influenza vaccination, age >65	✓		HCQI	HCQI participating countries	Not specified	Judged suitable for international comparison[75] Comparability issues: survey vs. administrative data
Incidence rate (per 100,000 population) measles		✓	HCQI	HCQI participating countries	Not specified	Judged suitable for international comparison[75] Variation among countries regarding reporting (mandatory vs. voluntary of confirmed/suspected cases); impact considered to be small
Incidence rate (per 100,000 population) pertussis		✓	HCQI	HCQI participating countries	Not specified	As above
Incidence rate (per 100,000 population) hepatitis B		✓	HCQI	HCQI participating countries	Not specified	As above
Cancer						
Mammography screening rate among women aged 50-69 within past year	✓		HCQI	HCQI participating countries	Not specified	Judged suitable for international comparison[75] Comparability issues: survey vs. administrative data; variation of eligibility criteria
Participation in breast screening programme, women aged 50-75	✓		*DHCPR 2008[42]*	*National data*	*Effectiveness of preventive care*	*EU guidelines for breast cancer diagnosis and screening used as benchmark*
5-year survival rate for breast cancer (observed and relative)		✓	HCQI	HCQI participating countries	Not specified *(RC)*ß	Judged suitable for international comparison[75] Comparability issues: see case study 'Cancer survival'
Cervical cancer screening rate among women aged 20-69 within past 3 years	✓		HCQI	HCQI participating countries	Not specified	Judged suitable for international comparison[75] Comparability issues: survey vs. administrative data; variation of eligibility criteria
5-year survival rate for cervical cancer (observed and relative)		✓	HCQI	HCQI participating countries	Not specified *(RC)*ß	Judged suitable for international comparison[75] Comparability issues: see case study 'Cancer survival'

Table 4 International quality indicators: effectiveness (by disease area) *(continued)*

Indicator	P§	O§	Used by	Data source	Indicator of	Comment
5-year survival rate for colorectal cancer (observed and relative)		✔	HCQI	HCQI participating countries	Not specified *(RC)*ß	Judged suitable for international comparison[75] Comparability issues: see case study 'Cancer survival'
5-year survival rate for breast cancer, cervical cancer or colon cancer		✔	DHCPR 2008[42]	OECD health data	Effectiveness of curative care	Composite index developed by RIVM using OECD data (13 OECD countries)
Cardiovascular disease						
Patient/admission based in-hospital mortality rate within 30 days of hospital admission for acute myocardial infarction		✔	HCQI	HCQI participating countries	Not specified *(RC)*ß	Judged suitable for international comparison[75]
Patient/admission based in-hospital mortality rate within 30 days of hospital admission for hemorrhagic stroke ¥		✔	HCQI	HCQI participating countries	Not specified *(RC)*ß	Judged suitable for international comparison (stroke combined)[75]
Patient/admission based in-hospital mortality rate within 30 days of hospital admission for ischaemic stroke ¥		✔	HCQI	HCQI participating countries	Not specified *(RC)*ß	As above
30-day hospital mortality rate for acute myocardial infarction, cerebral haemorrhage and cerebral infarction		✔	DHCPR 2008[42]	OECD health data	Effectiveness of curative care	Composite index developed by RIVM using OECD data (15 OECD countries)
Admission rate congestive heart failure (CHF) (age 15+) ¥	✔		HCQI	HCQI participating countries	Health promotion, prevention and primary care	Comparability issues not specified Included in 2008-09 data collection
Admission rate hypertension (age 15+)	✔		HCQI	HCQI participating countries	Health promotion, prevention and primary care	Judged NOT suitable for international comparison (2007)[36] although included in 2008-09 data collection Comparability issues concern variation in hypertension prevalence rates across countries and in the definition of hypertension admissions

Table 4 International quality indicators: effectiveness (by disease area) *(continued)*

Indicator	P[§]	O[§]	Used by	Data source	Indicator of	Comment
Admission rate angina without procedure (age 15+) ¥	✓		HCQI	HCQI participating countries	Health promotion, prevention and primary care	Comparability issues not specified Included in 2008-09 data collection
Respiratory disease						
Asthma mortality ages 5-39		✓	HCQI	HCQI participating countries	Not specified *(RC)*[ß]	Judged suitable for international comparison[75]
			DHCPR 2008[42]	OECD health data	Effectiveness of curative care	
Asthma admission rate (age 15+)	✓		HCQI	HCQI participating countries	Health promotion, prevention and primary care	Judged suitable for international comparison[36]
COPD admission rate (age 15+) ¥	✓		HCQI	HCQI participating countries	Health promotion, prevention and primary care	Comparability issues not specified Included in 2008-09 data collection
Diabetes						
Annual retinal exam for diabetics	✓		HCQI	HCQI participating countries	Not specified (RC)[ß]	Judged suitable for international comparison[36] Comparability issues concern differences in data sources (population-based surveys vs. clinical records); in 2007 only 10 countries could provide data
Annual HbA1c test for diabetics	✓		HCQI	HCQI participating countries	Not specified	Judged NOT suitable for international comparison (2007);[36] not included in 2008-09 data collection Comparability issues concern low number of countries that can provide relevant data; differences in data sources (survey vs. patient records); data often extracted from research project and so not generalisable
Poor glucose control (as defined by level of HbA1c)	✓		HCQI	HCQI participating countries	Not specified	Judged NOT suitable for international comparison (2007);[36] not included in 2008-09 data collection Comparability issues concern low number of countries that can provide relevant data; differences in definitions of 'poor glucose control'; data from patient records not easily available routinely[36]
Diabetes short-term complications admission rate (age 15+) ¥	✓		HCQI	HCQI participating countries	Health promotion, prevention and primary care	Comparability issues not specified Included in 2008-09 data collection

Table 4 International quality indicators: effectiveness (by disease area) *(continued)*

Indicator	P[§]	O[§]	Used by	Data source	Indicator of	Comment
Diabetes long-term complications admission rate (age 15+) ¥	✓		HCQI	HCQI participating countries	Health promotion, prevention and primary care	Comparability issues not specified Included in 2008-09 data collection
Uncontrolled diabetes admission rate (age 15+)	✓		HCQI	HCQI participating countries	Health promotion, prevention and primary care	Judged NOT suitable for international comparison (2007)[36] although included in 2008-09 data collection Comparability issues concern variation among countries as to the definition of 'uncontrolled diabetes admission' (numerator estimation)
Diabetes lower extremity amputation rate		✓	HCQI	HCQI participating countries	Health promotion, prevention and primary care	Judged NOT suitable for international comparison (2007)[36] although included in 2008-09 data collection Comparability issues concern variation among countries as to which procedures are being included; difficulties relating to estimating the denominator
Mental health						
Unplanned schizophrenia re-admission rate (age 15+)	✓		HCQI	HCQI participating countries	Mental healthcare	Indicator judged as meeting the core selection criteria (importance and scientific soundness[86] as well as availability in at least 12 countries[82])
Unplanned bipolar disorder re-admission rate (age 15+)	✓		HCQI	HCQI participating countries	Mental healthcare	As above
% adults with a severe anxiety, mood or addiction disorder who receive care for this	✓		DHCPR 2008[42]	World Mental Health Survey	Effectiveness of mental healthcare	Selected EU countries only (no data on UK); reflects data collection 2001-2003
% adults with a severe anxiety, mood or addiction disorder under care who receive at least one follow-up contact	✓		DHCPR 2008[42]	World Mental Health Survey	Effectiveness of mental healthcare	As above
% adults with a severe anxiety, mood or addiction disorder under care who receive a satisfactory form of care	✓		DHCPR 2008[42]	World Mental Health Survey	Effectiveness of mental healthcare	As above

Table 4 International quality indicators: effectiveness (by disease area) *(continued)*

Indicator	P§	O§	Used by	Data source	Indicator of	Comment
Surgery						
Hip fractures that are operated on within 48 hours (%)	✓		DHCPR 2008[42]	OECD health data	Effectiveness of curative care	
In-hospital waiting time for surgery after hip fracture, age >65	✓		HCQI	HCQI participating countries	Not specified *(RC)*ß	Judged suitable for international comparison[75]
Other						
Infant mortality per 1,000 live births		✓	DHCPR 2008[42]	OECD health data	Effectiveness of prevention	Infant mortality is interpreted as providing "an indication of the effects of the economic and social situation on the health of mothers and newborn babies" (DHCPR 2008:21)
			US National Scorecard	OECD health data	Health outcomes	
Child survival (under 5 years)		✓	WHR 2000	World Health Organization	Health improvement goal	Calculated as index of equality to indicate distribution of health in a given system
Mortality amenable to healthcare per 100,000 population		✓	US National Scorecard	Nolte and McKee (2003, 2008)	Health outcomes	Measure "gauge[s] the extent to which health care services save lives and contribute to longer population life"[83]
Healthy life expectancy at age 60, by sex		✓	US National Scorecard	World Health Report 2003	Health outcomes	Aggregate measure that reflects range of factors outside health system such as living and working conditions; however, timely and effective healthcare can prevent/delay onset of disabling health conditions, while improving functioning and reducing complications from such illnesses[83]
Health attainment (disability-adjusted life expectancy [DALE])		✓	WHR 2000	World Health Organization	Health improvement goal	
% patients having received lifestyle counselling or coaching (weight, nutrition and exercise) from GPs during the past 2 years	✓		DHCPR 2008[42]	Commonwealth Fund Survey	Effectiveness of prevention	Origins of data unclear; source quoted in DHCPR does not report these data
Smoking prevalence		✓	HCQI	HCQI participating countries	Not specified	Judged suitable for international comparison[75]

§ P = process indicator; O = outcome indicator

¥ listed in OECD Health Care Quality Indicators Data Collection for 2008-09 / Guidelines for completing data questionnaires only

ß RC = Regular Collection

Table 5 International quality indicators: patient safety

Indicator	P[§]	O[§]	Used by	Data source	Indicator of	Comment
Patient reported medical, medication, and lab test error (% reporting)	✓		DHCPR 2008[42]	Commonwealth Fund Survey[61]	Patient safety	Based on survey data; lack of detail on questionnaire design, data quality, validity and reliability; comparability issues uncertain[78]
			US National Scorecard	Commonwealth Fund Survey[61 63]	Quality of care	Uses data from 2005 and 2007 survey to present time trends as above
Infection due to medical care		✓	HCQI	HCQI participating countries	Hospital-acquired infections	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] selected for further analysis of international data comparability[81] Not included in 2008-09 data collection
Decubitus ulcer		✓	HCQI	HCQI participating countries	Hospital-acquired infections	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] selected for further analysis of international data comparability[81] Not included in 2008-09 data collection
Complications of anaesthesia		✓	HCQI	HCQI participating countries	Operative and post-operative complications	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] selected for further analysis of international data comparability[81] Not included in 2008-09 data collection
Postoperative hip fracture		✓	HCQI	HCQI participating countries	Operative and post-operative complications	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] judged NOT suitable for international comparison (2007)[36] pending further analytical work and selected for further analysis of international data comparability[81] Not included in 2008-09 data collection
Post-operative pulmonary embolism or deep vein thrombosis		✓	HCQI	HCQI participating countries	Operative and post-operative complications	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] selected for further analysis of international data comparability[81] Included in 2008-09 data collection
Post-operative sepsis		✓	HCQI	HCQI participating countries	Operative and post-operative complications	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] selected for further analysis of international data comparability[81] Included in 2008-09 data collection

Table 5 International quality indicators: patient safety *(continued)*

Indicator	P[§]	O[§]	Used by	Data source	Indicator of	Comment
Technical difficulty with procedure (e.g. accidental puncture or laceration)		✓	HCQI	HCQI participating countries	Operative and post-operative complications	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] selected for further analysis of international data comparability[81] 'Accidental puncture/laceration' included in 2008-09 data collection
Post-operative respiratory failure		✓	HCQI	HCQI participating countries	Operative and post-operative complications	Selected for further analysis of international data comparability following review by the OECD Expert Group on Patient Safety[81] Not included in 2008-09 data collection
Iatrogenic pneumothorax		✓	HCQI	HCQI participating countries	Operative and post-operative complications	Selected for further analysis of international data comparability following review by the OECD Expert Group on Patient Safety[81] Not included in 2008-09 data collection
Transfusion reaction		✓	HCQI	HCQI participating countries	Sentinel events	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] judged NOT suitable for international comparison (2007)[36] pending further analytical work; selected for further analysis of international comparability[81] Not included in 2008-09 data collection
Foreign body left in during procedure		✓	HCQI	HCQI participating countries	Sentinel events	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] selected for further analysis of international data comparability[81] Included in 2008-09 data collection
Vascular catheter related infections ¥		✓	HCQI	HCQI participating countries	Not specified	Comparability issues not specified Included in 2008-09 data collection
Birth trauma, injury to neonate		✓	HCQI	HCQI participating countries	Operative and post-operative complications	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] judged NOT suitable for international comparison (2007)[36] pending further analytical work; selected for further analysis of international comparability[81] Not included in 2008-09 data collection
Obstetric trauma, vaginal delivery with instrument		✓	HCQI	HCQI participating countries	Operative and post-operative complications	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] selected for further analysis of international data comparability[81] Included in 2008-09 data collection

Table 5 International quality indicators: patient safety *(continued)*

Indicator	P§	O§	Used by	Data source	Indicator of	Comment
Obstetric trauma, vaginal delivery without instrument		✓	HCQI	HCQI participating countries	Operative and post-operative complications	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] selected for further analysis of international data comparability[81] Included in 2008-09 data collection
Obstetric trauma, caesarian section		✓	HCQI	HCQI participating countries	Operative and post-operative complications	Recommended for further consideration as HCQI PSI on basis of meeting core criteria for quality indicators (importance and scientific soundness);[87] judged NOT suitable for international comparison (2007)[36] pending further analytical work; selected for further analysis of international comparability[81] Not included in 2008-09 data collection

P§ = process indicator; O§ = outcome indicator

¥ listed in OECD Health Care Quality Indicators Data Collection for 2008-09 / Guidelines for completing data questionnaires only

HCQI PSI = OECD Health Care Quality Indicator project Patient Safety Indicators

Table 6 International quality indicators: patient experience

Indicator	Used by	Data source	Indicator of	Comment
Health system responsiveness	World Health Report 2000	World Health Organization	Health system responsiveness	Based on a survey of ~2,000 key informants in 35 countries and assessed as (combined) score of seven elements of responsiveness (*respect for persons*: dignity, autonomy and confidentiality; *client orientation*: prompt attention, quality of basic amenities, access to social support networks and provider choice); responsiveness levels estimated for countries not covered by survey using indirect techniques[33]
	DHCPR 2008[42]	World Health Survey 2002	Connecting theme	
Public views on the functioning of the healthcare system (minor change needed; fundamental change needed; rebuild completely) (% respondents age 18+)	DHCPR 2008[42]	Commonwealth Fund Survey[61]	Effectiveness of curative care	Based on survey data; lack of detail on questionnaire design, data quality, validity and reliability; comparability issues uncertain[78]
Confidence to receive high quality and safe care, the most effective drugs and/or the best medical technology (% respondents age 18+)	DHCPR 2008[42]	Commonwealth Fund Survey[61]	Effectiveness of curative care	As above
Looking for and finding useful information on quality information on doctors and hospitals (% respondents age 18+)	DHCPR 2008[42]	Commonwealth Fund Survey[61]	Patient choice	As above
Access to care				
Confidence in the affordability of necessary healthcare (% respondents age 18+)	DHCPR 2008[42]	Commonwealth Fund Survey[61]	Financial access to care	As above
% foregoing necessary healthcare because of costs	DHCPR 2008[42]	Commonwealth Fund Survey[61]	Financial access to care	As above
Speed of access to primary care doctor (days waiting for appointment when in need of care) (% respondents age 18+)	DHCPR 2008[42]	Commonwealth Fund Survey[61]	Waiting for regular care	As above
	US National Scorecard	Commonwealth Fund Survey[61]	Quality of care	As above
Access to care out of hours (easy – difficult) (% respondents age 18+)	DHCPR 2008[42]	Commonwealth Fund Survey[61]	Waiting for regular care	As above
	US National Scorecard	Commonwealth Fund Survey[61]	Quality of care	As above

Table 6 International quality indicators: patient experience *(continued)*

Indicator	Used by	Data source	Indicator of	Comment
Care coordination				
Patient has regular GP/doctor GP knows about important information about the patient's medical history GP/member of staff helps with coordinating their care (% respondents age 18+)	DHCPR 2008[42]	Commonwealth Fund Survey[61]	Connecting theme	As above
% adults reporting that doctor ordered test that had already been done in past 2 years	US National Scorecard	Commonwealth Fund Survey[61 63]	Efficiency	As above
% adults reporting test results/records not available at time of appointment in past 2 years	US National Scorecard	Commonwealth Fund Survey[61 63]	Efficiency	As above
% adults who went to Emergency Room in past 2 years for condition that could have been treated by regular doctor if available	US National Scorecard	Commonwealth Fund Survey	Efficiency	As above
Coordination of medication use for people admitted to hospital (medication review/discussion at admission/discharge)	DHCPR 2008[42]	Commonwealth Fund Survey[61]	Effectiveness of curative care	As above
% hospitalised patients with new prescription who reported prior medications were reviewed at discharge	US National Scorecard	Commonwealth Fund Survey[63]	Quality of care	As above
% adults with chronic conditions given self-management plan	US National Scorecard	Commonwealth Fund Survey[61 63]	Quality of care	As above
Care is personalised: GP spends enough time, explains, involves patient in decision-making	DHCPR 2008[42]	Commonwealth Fund Survey[61]	Connecting theme	As above

Case study (1): 'Avoidable' mortality

The concept of 'avoidable' mortality, as used over the last three decades, originates from the Working Group on Preventable and Manageable Diseases led by David Rustein of Harvard Medical School in the 1970s.[88] They introduced the notion of 'unnecessary untimely deaths' by proposing a list of conditions from which death should not occur in the presence of timely and effective medical care. These 'sentinel health events' were to serve as an index of the quality of care and so enabling the monitoring of healthcare quality over time and between systems.

The concept was subsequently adopted widely as a means to assess the quality of healthcare, particularly in Europe, applying it to routinely collected data. It gained momentum with the European Commission Concerted Action Project on Health Services and 'Avoidable Deaths', established in the early 1980s and leading to the publication of the European Community Atlas of 'Avoidable Death' in 1988,[89] a major work that has been updated twice.[90-92] However, only recently has the concept been revitalised as a potentially useful tool to assess the performance of health systems[6] and it has since been taken up widely as a means to measure healthcare quality nationally[83 93 94] and internationally.[69 71 95 96]

Figure 1 shows an example of the application of the concept, comparing the performance of 19 OECD countries on the indicator.[96] It finds that all countries experienced a fall in 'avoidable' mortality between 1997/98 and 2002/03, although the scale and pace of change varied. The largest reductions were seen in countries with the highest initial levels, including Portugal, Finland, Ireland and the UK, with rates falling by around 20%, but also in some countries that had been performing better initially, such as Italy and France, where rates fell by about 15%. In contrast, the USA also started from a relatively high level of avoidable mortality but experienced much smaller reductions, with rates falling by 4% only. This suggests that the UK has made considerable progress in reducing the mortality of conditions considered amenable to healthcare, and, by extension, in improving healthcare performance overall although levels of 'avoidable' mortality remain higher compared to other advanced healthcare systems.

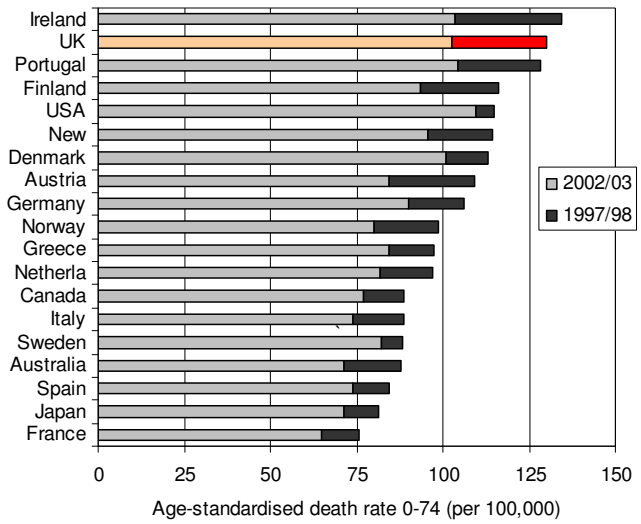

Figure 1 Mortality from conditions considered amenable to healthcare, age 0-74 years, in 19 OECD countries, 1997/98 and 2002/03 (Denmark: 2000/01; Sweden: 2001/02; Italy, USA: 2002)
***Source*: adapted from [4 96]**

However, while 'avoidable' mortality provides a valuable indicator of overall healthcare system performance, it does have several limitations. These have been discussed in detail.[6] Drawing on Nolte et al. (2009), we here highlight selected aspects that need to be considered when using the indicator 'avoidable' mortality to assess healthcare quality at the system level.

Thus one key limitation of the concept is its focus on mortality, which captures only part of the system and is not relevant for those healthcare services that are focused primarily on relieving pain and improving the quality of life. However, reliable data on morbidity that would allow for cross-national comparisons are still scarce. Countries are increasingly establishing disease registries in addition to cancer registries, for example for conditions such as diabetes, myocardial infarction or stroke; however, information may be misleading where registration focuses on selected population groups only, as is indeed the case for some cancer registries as outlined below. Population surveys provide another potential source of data on morbidity although, as noted earlier, survey data are often not comparable across regions. Routinely collected health service utilisation data such as hospital admissions data, as is currently explored by the OECD within their Health Care Quality Indicators project (Table 4, Table 5), or consultations of general practitioners and/or specialists, usually cover an entire region or country. Yet, while potentially useful, utilisation data, especially consultation rates, only include those segments of a population that seeks care but not those who may be in need of care but do not seek it.

A further limitation of the concept is its aggregate nature. A single figure of 'avoidable' mortality for an entire country, as for example shown in Figure 1, will inevitably conceal large variations at the sub-regional and sub-group level, for example as it relates to

geography, ethnicity, or access to care, among many other factors. Interpretation of the data must go beyond the aggregate figure to look within populations and at specific causes of death if these findings are to inform policy.

At a practical level, application of the concept for international benchmarking remains currently limited. Although it rests on routinely collected mortality data that are available for most countries, actual rates of 'avoidable' mortality are not routinely calculated at an international or at the national level. For England, the National Centre for Health Outcomes Development (NCHOD) routinely publishes data on 'avoidable' mortality (using the notion of 'amenable' mortality, see below) and these are currently available for the period 1993-2007 (Figure 2). However, NCHOD uses a slightly different definition of 'avoidable' mortality than the one illustrated in Figure 1;[2] therefore any direct comparison with data from other countries will carefully have to consider the definition of 'avoidable' mortality being used.

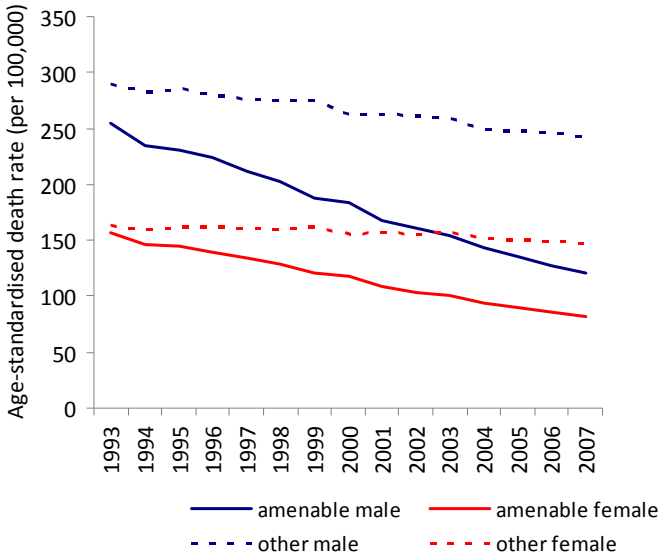

Figure 2 Age-standardised death rate (< 75 years), per 100,000 population, conditions considered amenable to healthcare and remaining conditions ('other'), by sex, England, 1993-2007
***Source*: adapted from [97]**

This last point highlights an important challenge related to any application of the concept of 'avoidable' mortality, namely the selection of conditions to be considered 'avoidable' given that a death from any cause is typically the final event in a complex chain of

[2] NCHOD includes all deaths (under age 75) from ischaemic heart disease into the summary measure of 'avoidable' mortality. However, Figure 1 includes only 50% of IHD (Nolte and McKee 2009). As a consequence, rates of 'avoidable' mortality as provided by NCHOD will inevitably be higher than those that are shown in Figure 1.

processes that include underlying social and economic factors and lifestyles, as well as preventive and curative healthcare. Some authors have used the term 'amenable' mortality to refer specifically to deaths that can be averted through healthcare interventions such as primary and hospital care, as well as preventive services such as screening and immunisation.[6 89 98 99] This interpretation defines 'amenable' conditions as those from which it is reasonable to expect death to be averted even after the condition has developed; it would thus include conditions such as tuberculosis where, although the acquisition of disease is largely driven by socio-economic conditions, timely treatment is effective in preventing death. Conversely, conditions such as lung cancer or liver cirrhosis would not be considered 'amenable' as their aetiology is, to a considerable extent, related to lifestyle factors, in this case the use of alcohol and tobacco, and prevention of death from these conditions is less likely to be susceptible to healthcare intervention.

It is also important to consider that the concept of 'avoidable' mortality as commonly used applies to premature deaths, commonly capturing deaths up to the age of 65 years only[6] although more recent work has extended the upper age limit to 75 years.[95 96] In the UK, in 2005, mortality under the age of 75 years captured between 47% (women) and 52% (men) of all deaths; of these, 25% (men) and 32.5% (women) were attributable to deaths considered amenable to healthcare. Thus, given that about half of the mortality experience now occurs above age 75, and against increasing evidence for healthcare interventions benefiting people at older ages, the application of 'avoidable' mortality that uses the age cut-off at 75 years is likely to underestimate the 'true' impact of healthcare on changes in population health.

Finally, any application of the concept 'amenable' or 'avoidable' mortality as a means to assess the quality of care of a given healthcare system will have to consider whether an observed change over time, as for example shown in Figure 1, can indeed be attributed to healthcare, or whether the change is simply due to a secular trend through for example improving living and socioeconomic conditions and 'spontaneous' declines in the incidence of causes considered amenable. To account for such confounding, researchers have compared trends in 'avoidable' mortality with trends in mortality from all other conditions that are not considered amenable to healthcare intervention ('other causes').[96 100] This is further illustrated in Figure 2, which shows that in England, between 1993 and 2007, amenable mortality (including all IHD deaths < 75) fell by 53% among men and 48% among women. In contrast, mortality from all other conditions fell by respectively 17% and 10% only. These findings suggest that healthcare did indeed contribute to improving population health as measured by amenable mortality.

In summary, as Nolte et al. (2009) have noted, while recognising the limitations of the concept of 'avoidable' mortality, it provides a potentially useful indicator of healthcare quality at the system level. It is, however, important to stress that high levels of 'avoidable' mortality should not be mistaken as definitive evidence of ineffective healthcare but rather as an indicator of potential weaknesses that can then be investigated in more depth.

Case study (2): Cancer survival

Cancer outcomes, in particular population-based cancer survival, are widely used as one measure of overall progress in cancer control and the overall effectiveness of health services.[24 74 101 102] Cancer survival statistics have intrinsic appeal as a measure of overall system progress in cancer control as cancer is common, it causes a large proportion of total deaths, and it is one of the few diseases where individual survival data are often captured routinely in a readily accessible format. This has led to their widespread use for assessing differences within sub-groups in populations cross-sectionally[103] and over time,[104 105] with for example across Europe the Nordic countries generally showing the highest survival rates for most common cancers (Figure 3), and evidence of even more marked differences between Europe and the USA.[106]

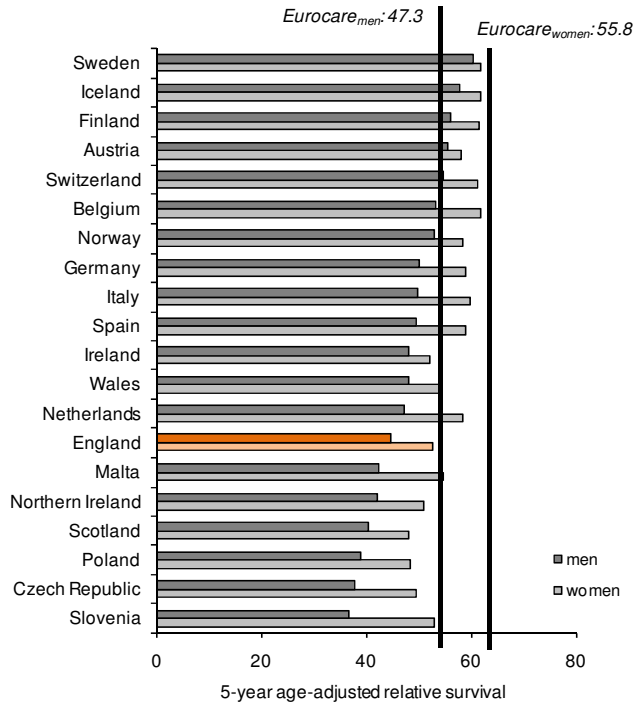

Figure 3 Age-adjusted 5-year relative survival of all malignancies of men and women diagnosed in 2000-02

***Source*: adapted from [4 107]**

However, commentators have highlighted the many elements that influence cancer outcomes.[103 106] Main concerns surround the notion of the 'case-mix', relating to issues around data completeness or coverage, i.e. not all registries cover the entire population and so do not necessarily represent the socio-demographic composition of the population in question; the distribution of tumour stages, which will depend on whether there are screening programmes, as with prostate and breast cancer, and other aspects of disease detection; and time lags (personal and system-induced) between symptom occurrence and treatment.[108] We will examine these concerns in turn.

Data completeness, population covered

The proportion of national populations covered by cancer registries that often provide the data for international comparative studies varies widely. For example, cancer data as captured in the EUROCARE study represent only just over 1% of the entire population in Germany and between 10% and 15% of the population in France compared with 100% in England, Denmark and Sweden.[109] This diversity in terms of population coverage will inevitably limit the interpretability of data for some regions in particular.[110]

Data from the USA suggest that the rather select nature of the populations covered by the registries of the SEER (Surveillance Epidemiology and End Results) programme, widely used in international comparisons, accounts for much of the seemingly better survival in the USA for a number of major cancers.[111] Indeed, as Mariotto et al. (2002) illustrated, adjustment of SEER rates to reflect the characteristics of the US population caused 5-year death rates to increase by 15% for prostate cancer, by 12% for breast cancer and by 6% for colorectal cancer in men, bringing them quite close to European survival figures.

Distribution of tumour stage, diagnosis

Cancer survival depends, to a large degree, on the availability of diagnostic means for early detection, such as through screening programmes, as well as of appropriate treatment modalities. However, the availability of screening methods for selected cancers introduces two forms of biases with implications for the interpretation and comparison of cancer survival estimates. Thus screening may allow for the diagnosis of a given cancer before it is clinically evident and may 'artificially' increase survival time (lead-time bias). A related issue is that with more advanced screening methods there is greater chance of finding indolent (i.e. slow growing) cancers. However, many indolent cancers are not necessarily life-threatening and are thus unlikely to affect mortality, as for example with prostate cancer. If screening was not used, many patients diagnosed with indolent cancer would not have been included in the cohort of patients for which survival is being calculated (and so inflating estimates) (length-time bias).

This latter point was recently illustrated using EUROCARE data on 5-year relative survival from prostate cancer in a number of European countries (diagnosed in 1995-1999), along with data on incidence and mortality for the years 1995 and 2000. In doing so, Autier et al. (2007) demonstrated how the incidence of prostate cancer in 1995 was lower in the UK and in Denmark (combined) than in eight other countries, at 35.5/100,000 and

59.3/100,000, respectively; yet mean mortality in 2000 (i.e. after 5 years) was fairly similar, at 17.7/100,000 and 17.3/100,000.[112] These substantial differences in incidence against similar mortality rates likely explain the relatively worse outcomes for 5-year relative survival in both the UK and Denmark. However, based on the data, the authors further note that:

- cross-country variability in 5-year survival was highest for screen-detectable cancers such as prostate and breast cancer, and melanoma, and lowest for cancers for which there are no methods for early detection, such as ovarian and lung cancer;
- the 5-year relative survival rate was positively associated with incidence for cancers that are screen-detectable while there was no association with mortality; and
- for those cancers for which there are no screening methods, 5-year survival was (significantly and) negatively associated with mortality, as well as with incidence, reflecting the rapidly fatal outcome of these cancers.[112]

These findings suggest that an observed longer survival noted for prostate and breast cancer as well as melanoma may be in part attributable to 'overdiagnosis' rather than be explained by improvement in the effectiveness of healthcare per se.

Also, there is sometimes uncertainty about the diagnosis of malignancy.[113] For example, there is some suggestion that apparently dramatic improvements in survival among US women with ovarian cancer in the late 1980s may be attributable, to a large extent, to changes in the classification of borderline ovarian tumours.[114] The ongoing CONCORD study is examining these issues in detail across four continents, supporting future calibration and interpretation of cancer survival rates.[106 115]

Time lags

Discussing the UK's relatively poor performance in relation to cancer survival as assessed through the EUROCARE study, Anderson and Murtagh (2007) argued that this may in part reflect patient characteristics, citing evidence of later presentation of patients compared with other countries,[110] with cancers diagnosed at a later stage as shown for colorectal cancer in England (Thames region) in the early 1990s.[116 117] Later diagnosis in European countries has also been identified as one factor explaining the apparently higher survival rates from a range of common cancers reported for the USA.[108]

Should cancer survival rates be used in international comparison of healthcare quality?

A key challenge to using cancer outcome data such as cancer survival as a measure of healthcare quality is that common to using outcome measures generally, namely the difficulty of attributing observed changes in the outcome of interest to changes in the healthcare system. Progress in treatment is rarely followed by rapid change in population-level long-term survival rates, mainly because of the time lag between the introduction of an intervention and its observable impact at the population level due to the usually incremental process with which a given protocol and/or interventions will become

accessible to increasing parts of the population.[52] An exception has been the introduction of cisplatin for the treatment of testicular cancer in the 1970s, the impact of which was observable almost immediately at the population level, most dramatically perhaps illustrated by the marked reduction in deaths from testicular cancer in the former German Democratic Republic (GDR) when modern chemotherapeutic agents became available after unification with west Germany.[118]

Cancer survival is a complex indicator and longer survival may reflect earlier diagnosis, over-diagnosis or later death.[52] However, it is important to note that registry data that currently form the basis of such estimates do not allow disentangling the various factors contributing to longer survival.[116] At present, the only adjustments incorporated in routine survival data are for age and the underlying general mortality rate of a population. Use of stage-specific rates would improve comparability,[115] but this is not widely available. However, even this is imperfect for comparisons of health systems at different 'evolutionary' stages: a more sophisticated staging system based on intensive diagnostic workup can improve stage-specific survival for all stages, as those transferred from the lower stage will usually have lower survival than those remaining in the former group, but better survival than those initially in the higher stage.

Based on the above there is perhaps little doubt that survival rates should at present be considered as no more than a means to flag initial possible concerns about the quality of care delivered in a given context.[4] However, some commentators argue that survival data should not be used for comparative purpose. Thus Autier et al. (2007) warn that, since relative survival data are susceptible to biases and limitations of incidence and mortality data, international comparisons should use mortality data as the most reliable indicator for cross-national comparisons of cancer outcomes only (also since mortality data normally cover the entire population of a given country),[112] although others have argued that using mortality data on their own is unlikely to offer insights superior to those offered by survival data.[116]

Dickman and Adami (2006) recently noted that, "in order to evaluate progress against cancer one must simultaneously interpret trends in incidence, mortality and survival" as none of the three measures is fully interpretable without knowledge of the other two.[119] This point is illustrated by a recent study by Sant et al. (2006), who examined breast cancer survival trends across Europe, using data on survival, incidence and mortality.[120] Looking at the UK, among other countries, it found a rise in survival throughout the 1980s and 1990s against a marked decrease in mortality and a tendency to incidence stabilisation during the past 30 years, pointing to the combined impact of early diagnosis through screening and treatment from around the late 1980s, a point that has also been made elsewhere against the very rapid decline in mortality from breast cancer in the UK since 1990 (Figure 4), pointing to the impact of improvements in early diagnostics and treatment.[121]

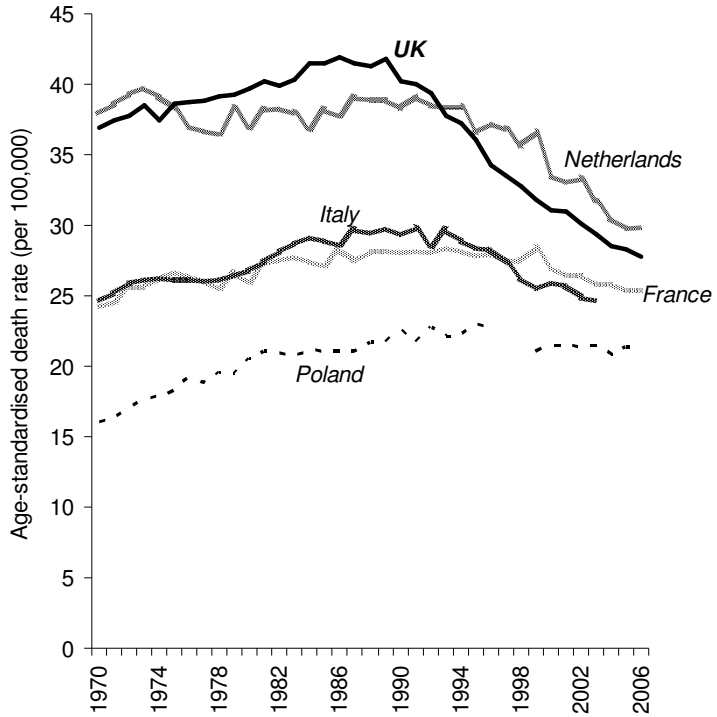

Figure 4 Age-standardised death rates from breast cancer in five countries, 1970-2006
***Source*: adapted from [67]**

Thus a detailed assessment of progress of a given system optimally would include a 'parallel' approach involving cross-sectional and longitudinal analyses, which, in the case of cancer, should, along with incidence and mortality data, ideally include stage-specific survival data so as to account for potential biases inherent in using short-term survival to assess screening effects.

Conclusions

There is growing interest in the systematic assessment and international benchmarking of the quality of care provided by healthcare systems, and major developments are under way to support this process through the development and validation of quality indicators, the systematic analysis of the suitability of existing datasets and the development and implementation of standard definitions and algorithms to improve the comparability of national data systems. Thus although cross-national data comparability remains a challenge, there is now a considerable body of evidence and actual data that allow for cross-national comparison of healthcare quality in selected areas of care.

However, while international comparison may provide an important benchmark for national progress, it will be important to consider using a range of indicators to capture the different aspects of a given aspect of healthcare in order to allow meaningful interpretation of observed phenomena, as in the example of cancer survival.

This report has focused on three quality domains only: effectiveness of care, patient safety and patient experience. However, it is important to recognise that patient outcomes are affected by a range of factors, and access to healthcare is an important additional component of quality, which may be a key determinant of differences in outcome between different countries. Indeed, access to care is considered an important domain of quality in several frameworks, including the WHO's health systems framework, the Dutch Health Care Performance framework and the US National Scorecard. Thus comparing quality in different countries is only a first step to subsequently assess the reasons for those differences, thereby determining what actions may be appropriate to take to improve health outcomes.

International benchmarking of quality of care with the NHS has considerable potential to improve patient outcomes in England, but only if research is undertaken to understand the reasons for differences between countries and their possible relevance to England.

List of references

1. Department of Health. High quality care for all. NHS next stage review final report. London: The Stationery Office, 2008.
2. Legido-Quigley H, McKee M, Nolte E, Glinos I. *Assuring the quality of health care in the European Union*. Copenhagen: World Health Organization on behalf of the European Observatory on Health Systems and Policies, 2008.
3. Nolte E, Bain C, McKee M. Chronic diseases as tracer conditions in international benchmarking of health systems: the example of diabetes. *Diabetes Care* 2006;29:1007-11.
4. Nolte E, Bain C, McKee M. Population health. In: Smith P, Mossialos E, Leatherman S, Papanicolas I, editors. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge: Cambridge University Press, 2009.
5. Nolte E, Ettelt S, Thomson S, Mays N. Learning from other countries: an on-call facility for health care policy. *J Health Serv Res Policy* 2008;13(Suppl. 2):58-64.
6. Nolte E, McKee M. *Does healthcare save lives? Avoidable mortality revisited*. London: The Nuffield Trust, 2004.
7. Nolte E, McKee M, Wait S. Describing and evaluating health systems. In: Bowling A, Ebrahim S, editors. *Handbook of health research methods: investigation, measurement and analysis*. Maidenhead: Open University Press, 2005:12-43.
8. Campbell S, Braspenning J, Hutchinson A, Marshall M. Research methods used in developing and applying quality indicators in primary care. *BMJ* 2003;326:816-19.
9. Smith P, Mossialos E, Papanicolas I. Performance measurement for health system improvement: experiences, challenges and prospects. Copenhagen: World Health Organization on behalf of the European Observatory on Health Systems and Policies, 2008.
10. Pringle M, Wilson T, Grol R. Measuring "goodness" in individuals and healthcare systems. *BMJ* 2002;325:704-7.
11. Davies H. Measuring and reporting the quality of health care: issues and evidence from the international research literature. Edinburgh: NHS Quality Improvement Scotland, 2005.
12. Freeman T. Use of performance indicators to improve health care quality in the public sector: a review of the literature. *Health Serv Manage Res* 2002;15:126-37.

13. Leatherman S. Applying performance indicators to health system improvement. *Measuring up. Improving health system performance in OECD countries*. Paris: OECD, 2002:319-33.
14. McGlynn E. Selecting common measures of quality and system performance. *Medical Care* 2003;41(Suppl.):I-39-I-47.
15. McGlynn E. Measuring clinical quality and appropriateness. In: Smith P, Mossialos E, Leatherman S, Papanicolas I, editors. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge: Cambridge University Press, 2009.
16. Mant J. Process versus outcome indicators in the assessment of quality of health care. *Int J Qual Health Care* 2001;13:475-80.
17. Evans D, Tan-Torres Edejer T, Lauer J, Frenk J, Murray C. Measuring quality: from the system to the provider. *Int J Qual Health Care* 2001;13:439-46.
18. Powell A, Davies H, Thomson R. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Qual Saf Health Care* 2003;12:122-28.
19. Joint Commission on Accreditation of Healthcare Organizations. *Primer on indicators development and application*. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations, 1990.
20. Walshe K. International comparisons of the quality of health care: what do they tell us? *Quality and Safety in Health Care* 2003;12:4-5.
21. McGlynn E, Asch S, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003;348:2635-45.
22. OECD. The OECD Health Care Quality Indicators Project. Background paper for the first meeting of the Expert Group. Paris: OECD, 2003.
23. National Centre for Health Outcomes Development. Evaluating the quality of clinical and health indicators, 2005. [Online] Available at: http://www.nchod.nhs.uk/NCHOD/Method.nsf/d3034ecd39b65e4965256cd2001d834f/7617d49c31d947118025701f004f1e84!OpenDocument (Accessed on 24 March 2009.
24. Hussey PS, Anderson GF, Osborn R, Feek C, McLaughlin V, Millar J, et al. How does the quality of care compare in five countries? *Health Affairs* 2004;23:89-99.
25. Campbell S, Hann M, Roland M, Shekelle P, Quayle J. The effect of panel membership and feedback on ratings in a two round Delphi survey: results of a randomised controlled trial. *Medical Care* 1999;37:964-68.
26. Heuschmann P, Biegler M, Busse O, Elsner S, Grau A, Hasenbein U, et al. Development and implementation of evidence-based indicators for measuring quality of acute stroke care. *Stroke* 2006;37:2573-78.
27. Steel N, Melzer D, Shekelle P, Wenger N, D F, McWilliams B. Developing quality indicators for older adults: transfer from the USA to the UK is feasible. *Qual Saf Health Care* 2004;13:260-64.
28. van der Ploeg E, Depla M, Shekelle P, Rutger H, Mackenbach J. Developing quality indicators for general practice care for vulnerable elders; transfer from US to The Netherlands. *Qual Saf Health Care* 2008;17:291-95.

29. Giuffrida A, Gravelle H, Roland M. Measuring quality of care with routine data: avoiding confusion between performance indicators and health outcomes. *BMJ* 1999;319:94-98.
30. OECD. OECD Health Data 2007. Paris: OECD, 2007.
31. Shaw C, Kalo I. A background for national quality policies in health systems. Copenhagen: WHO Regional Office for Europe, 2002.
32. OECD. Measuring health care, 1960-1983 expenditure, costs and performance. Paris: OECD, 1985.
33. World Health Organization. The World Health Report 2000. Health systems: improving performance. Geneva: World Health Organization, 2000.
34. OECD. Measuring up. Improving health system performance in OECD countries. Paris: OECD, 2002.
35. Arah OA, Klazinga NS, Delnoij DMJ, Ten Asbroek AHA, Custers T. Conceptual framework for health systems performance: a quest for effectiveness, quality and improvement. *International Journal of Quality in Health Care* 2003;15:377-98.
36. Garcia Armesto S, Gil Lapetra M, Wei L, Kelley E, HCQI Expert Group. Health care quality indicators project 2006 data collection update report. *OECD Health Working Papers No. 29*. Paris: OECD, 2007.
37. NHS Executive. The NHS performance assessment framework. London: Department of Health, 1998.
38. Canadian Institute for Health Information, Statistics Canada. Canadian Health Information Roadmap Initiative Indicators Framework. Ottawa: Canadian Institute for Health Information, 2000.
39. National Health Performance Committee. National health performance framework report. Brisbane: NHPC, 2001.
40. Agency for Healthcare Research and Quality. Measuring healthcare quality, 2009. [Online] Available at: http://www.ahrq.gov/qual/measurix.htm (Accessed on 16 June 2009).
41. Mainz J, Randrup Krog B, Bjornshave B, Bartels P. Nationwide continuous quality improvement using clinical indicators: the Danish National Indicator Project. *Int J Qual Health Care* 2004;16(Suppl.1):i45-i50.
42. Westert G, van den Berg M, Koolman X, Verkleij H. Dutch health care performance report 2008. Bilthoven: National Institute for Public Health and the Environment, 2008.
43. Westert G, Verkleij H. Dutch health care performance report. Bilthoven: National Institute for Public Health and the Environment, 2006.
44. Heugren M, Aberg A, Koester M, Ljung R. Performance indicators in Swedish health care. *BMC Health Serv* 2007;7(Suppl):A10.
45. Swedish Association of Local Authorities and Regions, Swedish National Board of Health and Welfare. Quality and efficiency in Swedish health care - regional comparisons 2007. Stockholm: Swedish Association of Local Authorities and Regions/Swedish National Board of Health and Welfare, 2007.
46. Chiu W-T, Yang C-M, Lin H-W, Chu T-B. Development and implementation of a nationwide health care quality indicator system in Taiwan. *Int J Qual Health Care* 2007;19:21-28.

47. Canadian Institute for Health Information. About health indicators, 2003. [Online] Available at: http://www.cihi.ca/indicators/en/about.shtml (Accessed on 16 June 2009).
48. Canadian Institute for Health Information. Health indicators 2009. Ottawa: Canadian Institute for Health Information, 2009.
49. Roland M. Linking physician pay to quality of care: a major experiment in the UK. *N Engl J Med* 2004;351:1448-54.
50. Duckett S, Ward M. Developing 'robust performance benchmarks' for the next Australian Health Care Agreements: the need for a new framework. *Aust N Z J Health Policy* 2008;5:1.
51. National Health and Hospitals Reform Commission. Beyond the blame game: accountability and performance benchmarks for the next Australian Health Care Agreements. Canberra: National Health and Hospitals Reform Commission, 2008.
52. Berrino F, Verdecchia A, Lutz J, Lombardo C, Micheli A, Capocaccia R. Comparative cancer survival information in Europe. *Eur J Cancer* 2009;45:901-08.
53. EUROCARE. EUROCARE Survival of cancer patients in Europe, 2009. [Online] Available at: http://www.eurocare.it/ (Accessed 16 June 2009).
54. EUCID. Final report European Core Indicators in Diabetes project, 2008. [Online] Available at: http://ec.europa.eu/health/ph_projects/2005/action1/action1_2005_11_en.htm (accessed on 26 February 2009).
55. ECHIM. ECHIM, 2009. [Online] Available at: http://www.echim.org/index.html (Accessed on 26 February 2009).
56. Kilpelaeinen K, Aromaa A, Echim Core Group. European health indicators: development and initial implementation. Final report of the ECHIM project. Helsinki: National Public Health Institute, 2008.
57. Fédération nationale des observatoires régionaux de la santé. Health indicators in the European regions. Project ISARE. Paris: Fédération nationale des observatoires régionaux de la santé, 2007.
58. ISARE. ISARE, 2009. [Online] Available at: http://www.isare.org/ (Accessed on 26 February 2009).
59. Campbell S, Ludt S, Van Lieshout J, Boffin N, Wensing M, Petek D, et al. Quality indicators for the prevention and management of cardiovascular disease in primary care in nine European countries. *Eur J Cardiovasc Prev Rehab* 2008;15:509-15.
60. Engels Y, Dautzenberg M, Campbell S, Broge B, Boffin N, Marshall MN, et al. Testing a European set of indicators for the evaluation of the management of primary care practices. *Fam Pract* 2006;23:137-47.
61. Schoen C, Osborn R, Doty M, Bishop M, Peugh J, Murukutla N. Towards higher-performance health systems: adults' health care experiences in seven countries. *Health Affairs* 2007;26:w717-w34.
62. Schoen C, Osborn R, Huenh P, Doty M, Davis K, Zapert K, et al. Primary care and health system performance: adult's experiences in five countries. *Health Affairs* 2004;W4:487-503.

63. Schoen C, Osborn R, Huynh P, Doty M, Zapert K, Peugh J, et al. Taking the pulse of health care systems: experiences of patients with health problems in six countries. *Health Affairs* 2005;24:w509-w25.
64. Schoen C, Osborn R, How S, Doty M, Peugh J. In chronic condition: experiences of patients with complex health care needs, in eight countries, 2008. *Health Affairs* 2009;28:w1-16.
65. Valentine N, Prasad A, Rice N, Robone S, Chatterji S. Health systems responsiveness - a measure of the acceptability of health care processes and systems from the user's perspective. In: Smith P, Mossialos E, Leatherman S, Papanicolas I, editors. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge: Cambridge University Press, 2009.
66. World Health Organization. World health survey, 2009. [Online] Available at: http://www.who.int/healthinfo/survey/en/ (Accessed on 16 June 2009).
67. World Health Organization Regional Office for Europe. European health for all database. Updated January 2009. Copenhagen: World Health Organization Regional Office for Europe, 2009.
68. OECD. OECD Health Data 2008: statistics and indicators for 30 OECD Countries. Paris: OECD, 2008.
69. The Commonwealth Fund Commission on a High Performance Health System. Why not the best? Results from a national scorecard on U.S. health system performance. New York: The Commonwealth Fund, 2006.
70. The Commonwealth Fund Commission on a High Performance Health System. Framework for a high performance health system for the United States. New York: The Commonwealth Fund, 2006.
71. The Commonwealth Fund Commission on a High Performance Health System. Why not the best? Results from the national scorecard on U.S. health system performance. New York: The Commonwealth Fund, 2008.
72. Murray C, Frenk J. A framework for assessing the performance of health systems. *Bull World Health Organization* 2000;78:717-31.
73. Murray C, Frenk J. World Health Report 2000: a step towards evidence-based health policy. *Lancet* 2000;357:1698-700.
74. Kelley E, Hurst J. Health care quality indicators project. Conceptual framework paper. *OECD Health Working Papers 23*. Paris: OECD, 2006.
75. Mattke S, Kelley E, Scherer P, Hurst J, Gil Lapetra M, HCQI Expert Group Members. Health Care Quality Indicators Project: Initial indicators report. Paris: OECD, 2006.
76. Mattke S, Epstein A, Leatherman S. The OECD Health Care Quality Indicators Project: history and background. *Int J Qual Health Care* 2006;18:1-4.
77. OECD. Health at a glance 2007 - OECD indicators. Paris: OECD, 2007.
78. Garratt A, Solheim E, Danielsen K. National and cross-national surveys of patient experiences: a structured review. Oslo: Norwegian Knowledge Centre for the Health Services, 2008.
79. Murray C, Evans D. Health systems performance assessment: goals, framework and overview. In: Murray C, Evans S, editors. *Health Systems Performance Assessment. Debates, methods and empiricism*. Geneva: World Health Organization, 2002.

80. Idänpään-Heikkilä U. The Nordic Indicator Group Project. Presentation at the OECD Health Care Quality Indicators Meeting, 13-14 January, 2003, Paris, 2003.
81. Droesler S. Facilitating cross national comparisons of indicators for patient safety at the health system level in the OECD countries. Paris: OECD, 2008.
82. Garcia Armesto S, Medeiros H, Wei L. Information availability for measuring and comparing quality of mental health care across OECD countries. Paris: OECD, 2008.
83. Schoen C, How S. National scorecard on U.S. health system performance: Technical report. New York: The Commonwealth Fund, 2006.
84. Veillard J, Garcia Armesto S, Kadandale S, Klazinga NS. International health system comparisons: from measurement challlenge to management tool. In: Smith P, Mossialos E, Leatherman S, Papanicolas I, editors. *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge: Cambridge University Press, 2009.
85. Care Quality Commission. Patient surveys, 2009. [Online] Available at: http://www.cqc.org.uk/usingcareservices/healthcare/patientsurveys.cfm (Accessed on 16 June 2009).
86. Hermann R, Mattke S, Members of the OECD Health Care Panel. Selecting indicators for the quality of mental health care at the health systems level in OECD countries. Paris: OECD, 2004.
87. Millar J, Mattke S, Members of the OECD Patient Safety Panel. Selecting indicators for patient safety at the health systems level in OECD countries. Paris: OECD, 2004.
88. Rutstein D, Berenberg W, Chalmers T, Child C, Fishman A, Perrin E. Measuring the quality of medical care. *N Engl J Med* 1976;294:582-88.
89. Holland W. *European Community atlas of 'avoidable death'*. Oxford, 1988.
90. Holland W, editor. *European Community atlas of 'avoidable death'. 2nd edition, Vol. I. Commission of the European Communities Health Services Research Series No. 6*. Oxford: Oxford University Press, 1991.
91. Holland W, editor. *European Community atlas of 'avoidable death'. 2nd edition, Vol. II. Commission of the European Communities Health Services Research Series No. 9*. Oxford: Oxford University Press, 1993.
92. Holland W, editor. *European Community atlas of 'avoidable death' 1985-89*. Oxford: Oxford University Press, 1997.
93. Gispert R, Torne M, Aran Bares M. The efficiency of the health system in Spain. *Gac Sanit* 2006;20 (Suppl. 1):117-26.
94. Korda R, Butler J. Effect of healthcare on mortality: Trends in avoidable mortality in Australia and comparisons with Western Europe. *Public Health* 2006;120:95-105.
95. Nolte E, McKee M. Measuring the health of the nations: how much is attributable to health care? An analysis of mortality amenable to medical care. *British Medical Journal* 2003;327:1129-32.
96. Nolte E, McKee M. Measuring the health of nations: updating an earlier analysis. *Health Affairs* 2008;27:58-71.

97. National Centre for Health Outcomes Development/The Information centre for health and social care. Mortality from causes considered amenable to health care, 2008. [Online] Available at: http://www.nchod.nhs.uk (Accessed on 24 March 2009).
98. Holland W. The 'avoidable death' guide to Europe. *Health Policy* 1986;6:115-17.
99. Mackenbach J, Looman C, Kunst A, Habbema J, van der Maas P. Post-1950 mortality trends and medical care: gains in life expectancy due to declines in mortality from conditions amenable to medical interventions in the Netherlands. *Soc Sci Med* 1988;27:889-594.
100. Mackenbach J, Bouvier-Colle M, Jougla E. "Avoidable" mortality and health services: a review of aggregate data studies. *J Epidemiol Commun Health* 1990;44:160-11.
101. Clauser S. Use of cancer performance measures in population health: a macro-level perspective. *J Natl Cancer Inst* 2004;33:142-54.
102. Rachet B, Maringe C, Nur U, Quaresma M, Shah A, Woods L, et al. Population-based cancer survival trends in England and Wales up to 2007: an assessment of the NHS cancer plan for Englad. *Lancet Oncol* 2009;10:351-69.
103. Coleman M, Babb P, Damiecki P, Grosclaude P, Honjo S, Jones J, et al. Cancer survival trends in England and Wales 1971-1995: deprivation and NHS region. London: The Stationary Office, 1999.
104. Berrino F, De Angelis R, Sant M, Rosso S, Bielska-Lasota M, Coebergh J, et al. Survival for eight major cancers and all cancers combined for European adults diagnosed in 1995-99: results of the EUROCARE-4 study. *Lancet Oncol* 2007;8:773-83.
105. Berrino F, Gatta G, Sant M, Capocaccia R. The EUROCARE study of survival of cancer patients in Europe: aims, current status, strengths and weaknesses. *Eur J Cancer* 2001;37:673-77.
106. Gatta G, Capocaccia R, Coleman M, Ries L, Hakulinen T, Micheli A, et al. Toward a comparison of survival in American and European cancer patients. *Cancer* 2000;89:893-900.
107. Verdecchia A, Francisci S, Brenner H, Gatta G, Micheli A, Mangone L, et al. Recent cancer survival in Europe: a 2000-02 period analysis of EUROCARE-4 data. *Lancet Oncol* 2007;8:784-96.
108. Sant M, Allemani C, Berrino F, Coleman M, Aareleid T, Chaplain G, et al. Breast carcinoma survival in Europe and the United States. *Cancer* 2004;100:715-22.
109. De Angelis R, Francisci S, Baili P, Marchesi F, Roazzi P, Belot A, et al. The EUROCARE-4 database on cancer survival in Europe: data standardisation, quality control and methods of statistical analysis. *Eur J Cancer* 2009;45:909-30.
110. Anderson W, Murtagh C. Cancer survival statistics should be viewed with caution. *Lancet Oncol* 2007;8:1052-53.
111. Mariotto A, Capocaccia R, Verdecchia A, Micheli A, Feuer E, Pickle L, et al. Projecting SEER cancer survival rates to the US: an ecological regression approach. *Cancer Causes Control* 2002;13:101-11.
112. Autier P, Boniol M, Hery CM, E, Ferlay J. Cancer survival statistics should be viewed with caution. *Lancet Oncol* 2007;8:1050-52.

113. Butler C, Currie G, Anderson W. Do differences in data reporting contribute to variation in lung cancer survival? *J Natl Cancer Inst* 2005;97:1385.
114. Kricker A. Ovarian cancer in Australian women. Camperdown, NSW: National Breast Cancer Center, 2002.
115. Ciccolallo L, Capocaccia R, Coleman M, Berrino F, Coebergh J, Damhuis R, et al. Survival differences between European and US patients with colorectal cancer: role of stage at diagnosis and surgery. *Gut* 2005;54:268-73.
116. Berrino F, De Angelis R, Sant M. Cancer survival statistics should be viewed with caution. *Lancet Oncol* 2007;8:1053-54.
117. Gatta G, Capocaccia R, Sant M, Bell C, Coebergh J, Damhuis R, et al. Understanding variations in survival for colorectal cancer in Europe: a EUROCAEE high resolution study. *Gut* 2000;47:533-38.
118. Becker N, Boyle P. Decline in mortality from testicular cancer in West Germany after reunification. *Lancet* 1997;350:744.
119. Dickman P, Adami H. Interpreting trends in cancer patient survival. *J Intern Med* 2006;260:103-17.
120. Sant M, Francisci S, Capocaccia R, Verdecchia A, Allemani C, Berrino F. Time trends of breast cancer survival in Europe in relation to incidence and mortality. *Int J Cancer* 2006;119(416-422).
121. Kobayashi S. What caused the decline in breast cancer mortality in the United Kingdom? *Breast Cancer* 2004;11:156-59.