

# Measuring and reporting the quality of health care:

issues and evidence from the international research literature

A Discussion Paper prepared by:

**Huw Davies**

Professor of Health Care Policy & Management,  
School of Management, University of St Andrews,

in consultation with  
**The Scottish Research Network  
in Health Care Management**

**© NHS Quality Improvement Scotland 2006**

Published November 2005

You can copy or reproduce the information in this document for use within NHSScotland and for educational purposes. You must not make a profit using information in this document. Commercial organisations must get our written permission before reproducing this document.

**[www.nhshealthquality.org](http://www.nhshealthquality.org)**

# Measuring and reporting the quality of health care: issues and evidence from the international research literature

A **Discussion Paper** prepared by: Huw Davies, Professor of Health Care Policy & Management, School of Management, University of St Andrews, in consultation with **The Scottish Research Network in Health Care Management**\*

## Remit:

1. to inform NHS QIS strategy development around the production and use of routine data on health care quality;
2. to do this by developing a narrative review drawing on the extensive research literature on measuring and reporting health care quality;
3. to highlight in particular two core issues: the development of robust and interpretable metrics; and the creation of systems that encourage data usage;
4. to summarise and report on this literature in a user-accessible style designed to appeal to a broad array of stakeholders.

## Contents:

- 1: Headline Findings [single page]
- 2: Structured Summary [4 pages]
- 3: Main Discussion Paper [20 pages]
- 4: References [5 pages]

## Disclaimer:

The contents of this Discussion Paper reflect the views of the author only, and should not be taken to reflect the views of *NHS Quality Improvement Scotland*, *The Scottish Research Network in Health Care Management* or any other agencies or individuals who may have been consulted during its preparation.

30-May-05

---

\* The Scottish Research Network on Health Care Management (SRN in HCM) is a collective endeavour across the Universities of Aberdeen, Edinburgh, Glasgow and St Andrews that seeks to coordinate academic and research-based expertise on contemporary health care management issues.

# Measuring and reporting the quality of health care: issues and evidence from the international research literature

A **Discussion Paper** prepared by: Huw Davies, Professor of Health Care Policy & Management, School of Management, University of St Andrews, in consultation with **The Scottish Research Network in Health Care Management**

## Headline findings

- The absence of data on individual professional practice has been identified as a significant barrier to achieving greater physician engagement in quality improvement activities.
- Yet, in general, centralised systems of analysis and reporting of comparative clinical data have only been shown to have relatively minor effects on quality of care.
- This may reflect more an absence of evidence than evidence of absence: *thus any new system should also seek to contribute to an evidence base of its own effects.*
- Physicians have significant anxieties about the public sharing of individualised data on clinical performance, and there is evidence that open and public reporting of quality data can induce a range of unwanted and dysfunctional effects.
- The creation of a new national system of quality indicators is therefore both a significant opportunity to address an important barrier to quality improvement, and a challenge to do so in a way that does not antagonise, demoralise or induce unintended consequences.
- Any system that seeks to compare clinical indicators across providers will need to address two core challenges:
  - the problems inherent in producing robust and interpretable metrics; and
  - the challenge of creating systems that encourage data usage in programmes of quality improvement at service level.
- The overarching organisations within which physicians are located can often prove the most sensitive of recipients for CCI data. *These organisations together with their front-line clinical staff should be the prime targets for CCI data.*
- At the heart of these arguments is the need to emphasise critical engagement with those whose practice is being measured, *and to reassure that the data gathered and analysed will be interpreted in the context of learning rather than judgement.*

30-May-05

# Structured Summary: *Reporting the quality of health care*

## Introduction

Many countries have experimented with the production and analysis of data on the quality and performance of health services, and Scotland took an early lead in this respect.

Two particular challenges emerge from this experience:

- the problems inherent in producing robust and interpretable metrics; and
- the challenge of creating systems that encourage data usage to improve care quality.

## **Comparative clinical indicators (CCIs)**

Comparative clinical indicators (CCIs) are measures, derived from routine data sets, that relate to the process and outcomes of clinical care. Comparisons may be historical, between various provider units, or against benchmarks.

While a focus on health outcomes has intuitive appeal, there may be good reasons why process measures offer significant advantages. The balance of relative merits suggests that national schemes should make use of a judicious mix of both types of measure while being fully cognisant of the potential difficulties and drawbacks of each.

## **Users and uses of comparative clinical indicators**

Potential *users* of comparative clinical indicators include: clinicians, health service managers, policy makers, the media and even the public.

Potential *uses* of comparative clinical indicators include their employment as part of: local quality improvement activities; service redesign; performance management and accountability activities; policy change; and public/media debates.

As we move through this continuum of uses, the key users shift from those most closely concerned with health care practice itself to those more involved in its management or oversight. At the same time, the degree of engagement shifts from voluntary to more mandatory, as does the role of incentives, moving from intrinsic to extrinsic drivers.

At the heart of these shifts is a change in the intent of usage from one focused on *local learning* to one more concerned with *making judgements* (although, of course, such judgements might ultimately be used in various ways to alter service delivery for the enhancement of care).

The emphasis in this Discussion Paper will be on the creation of CCIs that encourage learning.

However, we acknowledge that CCIs may be taken up and used in ways that are not part of their core purpose or design intent. Media spotlights in particular may fall on published outputs, and there is the potential that CCIs may begin to get incorporated elsewhere, for example into more formal process of accountability review, or as part of performance management strategies.

## **Creating robust and interpretable metrics**

The logic underpinning CCIs is that that variation across indicators reflects important variations in the underlying delivery of health care. The way in which the indicators are developed, defined, operationalised and analysed will have a big impact on the credibility of this logic.

## **Metric selection and design**

Systematic methods for combining evidence and expert opinion in the development of appropriate CCIs have been developed. These approaches can help address both face validity (professional consensus) and content validity (by explicit inclusion of available evidence).

Unjustly neglected in many current discussions is the extent to which any set of CCIs also reflects the core priorities and concerns of patients and users. This needs to be better addressed.

Useful indicators do not always need to be developed from scratch – they can be adopted and adapted from those used elsewhere in the world.

### **Measurement properties**

Any measure selected must pass muster against a wide range of technical criteria, principally: *validity* (does the measure capture what it is supposed to?), *reliability* (how reproducible are the measures when taken at different times or in different circumstances?), and *sensitivity to change* (will the measure discriminate between good and poor quality care, and be able to detect perhaps small but worthwhile improvements?).

It can be hard to assess the extent to which these criteria are met, and any changes to the detail of by who and how data are collected at local level may damage these technical properties.

### **Units for comparative analysis**

Data can be aggregated at the level of the individual practitioner, clinical team, organisational unit, or region/nation.

Since much modern health care, especially of chronic disease, is now delivered through networks of professionals that cut across traditional organisational boundaries then new units of analysis will be needed which focus on catchment populations rather than organisations *per se*.

### **Case-mix, confounding and case-mix adjustment**

Comparisons across quality indicators need to take into account the degree to which different ‘units of assessment’ are dealing with similar patient groups (statistical case-mix adjustment).

Developing and testing adjustment schemes is complex and incomplete – even the best models linking patient characteristics to patient outcomes explain only a portion of any variation.

Different adjustment schemes may lead to different rankings across key indicators, and there is a danger that, once attention is focused on those aspects of case mix that are being adjusted, these will then be assessed and/or recorded differently (a process called ‘upstaging’).

### **Coping with chance variability**

Traditional statistical techniques (such as p-values or confidence intervals) can aid interpretation amongst the statistically literate, but are confusing to many and are open to significant pitfalls.

Apparent differences across indicators may not in fact represent real underlying differences but may be the result of random variation. Falsely labelling providers as ‘poor performers’ may have considerable deleterious effects and is not that rare (research suggests that up to three-quarters of providers labelled as ‘poorly performing’ may in fact be no worse than average).

Conversely, real and important differences may go undetected when measures all fall within statistical limits (false reassurance).

Sophisticated techniques that can help assess the play of chance (such as statistical process control charts and multi-level models) have yet to see widespread acceptance or application.

Assessing changes over time in indicators can also be difficult.

### **Data quality and completeness**

Even the most well-chosen, carefully operationalised, and rigorously risk-adjusted indicators will be undermined if the data sets on which they draw are suspect. Yet research has shown repeatedly that routine data sets are frequently incomplete and inaccurate.

Research has also shown that indicators created from different data sources (e.g. data from clinical versus administrative records) can tell divergent stories about the quality of care.

### **Encouraging effective data usage**

If national systems of CCIs are to have the beneficial impacts hoped for then they need to be carefully targeted and embedded in a supportive infrastructure.

### ***Evidence and audience: who uses comparative clinical indicators?***

In general, centralised systems of analysis and reporting of CCIs have only been shown to have relatively minor effects on quality of care, although this may reflect more an absence of evidence than evidence of absence.

The public generally, and ‘consumers’ of health care more specifically, while expressing interest in these data, do not appear to use them to any great extent when making decisions.

Referring physicians (including general practitioners) also seem to make little use of quality data.

Physicians are sometimes aware of indicators when these relate to their own practice but are often sceptical of the information content.

The overarching organisations within which physicians are located can often prove the most sensitive of recipients for CCI data. Thus these organisations together with their front-line clinical staff should be the prime targets for CCI data.

### ***The importance of credibility***

Studies of how clinicians respond to data suggest the importance of having credible indicators with ‘scientific respectability’, in tandem with an honest portrayal of data limitations.

Additional important features may be the means by which the data are displayed and communicated, the extent to which the data suggest local courses of action, and data presentation in ways that illustrate and quantify the potential gains from quality improvement activities.

### ***Paying attention to motivation and incentives***

If sharing CCIs is to bring about improvements in health care then the patterns of behaviour of targeted audiences must be changed. This in turn highlights the importance of both *motivation* (the internalised desire for action) and how this might be elicited i.e. the role of *incentives*.

Core drivers to motivation can be thought of as either intrinsic or extrinsic. *Intrinsic* drivers are those internalised aspects of values and cognition that motivate particular patterns of behaviour, while *extrinsic* drivers are those external factors that shape behaviour, such as financial rewards, or regulatory demands.

A diverse range of incentives might be employed to drive quality improvement work informed by CCIs, including both financial and non-financial levers.

The evidence is pretty clear that financial incentives and other extrinsic drivers can elicit substantial shifts in behaviour patterns. What is less clear is the extent to which an emphasis on the extrinsic may deplete or damage intrinsic drivers.

Thus there need to be sufficient incentives around CCIs to encourage motivation to pay attention, take seriously, and act upon data; but drivers not so strong that they upset important components of intrinsic motivation, or induce dysfunctional responses.

### ***Support with tools for further analysis and action***

A first step should be systematic enquiry into the information needs, data presentation preferences, and associated analytic support, that might be required by different audiences.

Any national programme sharing CCIs should be very clear about how the data are meant to be disseminated and cascaded within provider organisations.

Support for data usage at local level may include: a programme of workshops to prompt active discussion with stakeholders; the commissioning of expert analysis and interpretation; the development of web-based facilities that allow access to the raw data and analytic tools; partnerships with funding bodies and change agencies so that providers have access to supportive funding and relevant expertise; and strengthening the linkages between different programmes of work in NHS QIS.

Engagement with provider organisations will need to be both intensive and bespoke. Local quality improvement and clinical governance arrangements will differ widely within and between organisations, and engagement around CCIs should reflect this.

As NHS QIS plans for supporting users are clarified it is suggested that further review work of the extensive organisational change literature might help to ensure that any package of measures is properly underpinned by theory and evidence of service-level change.

### ***Avoiding dysfunctional consequences***

All policy and managerial initiatives have the potential for unwanted, unexpected and dysfunctional consequences, and CCIs are no different in this regard.

Potential dysfunctional consequences range from the undesirable but not actively destructive (myopia and convergence) to the out-right corrupting (misrepresentation and fraud).

There are a number of ways in which the potential for dysfunctional consequences may be reduced, but some of these are in conflict suggesting the need for trade-offs.

While it can be hard to assess the extent to which dysfunctional consequences are likely, and difficult to evaluate the extent to which they are occurring in practice, their presence or otherwise is something that needs regular attention and review.

### **Concluding remarks**

The absence of data on individual professional practice has been identified as a significant barrier to achieving greater physician engagement in quality improvement activities.

At the same time, physicians have significant anxieties about the public sharing of individualised data on clinical performance.

The creation of national systems of CCIs is therefore both a significant opportunity to address an important barrier to quality improvement, and a challenge to do so in a way that does not antagonise, demoralise or induce unintended consequences.

At the heart of these arguments is the need to emphasise critical engagement with those whose practice is being measured, and to reassure that the data gathered and analysed will be interpreted in the context of learning rather than judgement.

Because of the (to date) limited evidence of effectiveness of national systems of CCIs, any new system should also seek to contribute to an evidence base of its own effects.

*END*

## Discussion Paper – Main Text

*'The NHS has yet to harness the power of data... Many in the NHS often view data with indifference or as a necessary evil... not as a useful tool to improve quality of care for patients within the local context.'* [Leatherman & Sutherland, 2003, p179].

### Introduction

#### **Background**

Many countries have experimented with the production and analysis of data on the quality and performance of health services [Ibrahim, 2001; McLoughlin et al., 2001; Mannion & Davies, 2002a; Marshall et al., 2003; 2004]. Sometimes such data are produced in aggregate form on a national basis, but more often they are presented broken down by region, or by hospital or other health care provider 'unit'. However careful the commentary provided alongside such comparative reporting, once made public these data are frequently seen as offering a 'report card' on the quality of health service delivery [Marshall et al., 2003]. While many have debated the relative merits of keeping such data confidential (for example seen only by the clinical teams concerned) or anonymous (data freely available but with the unit identifier disguised), trends towards more open and transparent governance in public services, combined with the impacts of the Freedom of Information Act, have largely made such debates redundant: these data, once collected and analysed, will be brought into the public domain.

Scotland was among the forerunners in developing this work, with the Clinical Resource and Audit Group (now part of NHS QIS) producing comparative clinical indicators for Scotland dating back to the mid 1990s. Preparation of these indicator sets was professionally-led, with health service organisations as the target audience (although the information was also released into the public domain, it was done so in a deliberately low-key manner). Other countries too (most notably the United States) have invested in a wide variety of clinical data analysis and feedback schemes aimed at improving accountability, performance, market operation and service quality [Davies & Marshall, 1999; Marshall et al., 2003]. From this wealth of local and international experience (and some concomitant evaluative effort) has emerged a substantial literature exploring the potential and pitfalls of collecting, analysing and using comparative data on clinical quality [McLoughlin et al., 2001; Powell et al., 2003; Marshall et al., 2003; 2004]. In particular, two core challenges emerge: first, the problems inherent in producing robust and interpretable metrics; and second, the challenges of embedding such metrics in systems that are truly likely to encourage use of the data in programmes of quality improvement at service level. It is to these two key challenges that this Discussion Paper is addressed.

#### **Defining comparative clinical indicators (CCIs)**

Comparative clinical indicators (CCIs) are measures, usually derived from routine data sets and collated centrally (whether regionally or nationally), that relate to the process and outcomes of clinical care. While a focus on health outcomes or other forms of direct patient experience (such as patient satisfaction) has intuitive appeal, there may be good reasons why process measures offer significant advantages [Davies & Crombie 1995; Mant & Hicks, 1995]. First, process measures have obvious salience to health care professionals and may more readily indicate what aspects of

care need attention; second, they make fewer demands for complex risk-adjustment (see later); and third, they may highlight important shortfalls in the standard of care much sooner than measures of risk-adjusted outcome [Mant & Hicks, 1995; Davies & Crombie, 1997; Crombie & Davies, 1998]. A fuller exploration of the relative advantages and disadvantages of process and outcome measures is outlined in Tables 1 and 2. This balance of relative merits suggests that national schemes should make use of a judicious mix of both types of measure while being fully cognisant of the potential difficulties and drawbacks of each [Mannion & Davies, 2002a].

**Table 1: Comparison of the relative advantages and disadvantages of PROCESS indicators**

<i>Advantages</i>	<i>Disadvantages</i>
1. <i>Readily measured</i> : utilization of health technologies is often relatively easily measured without major bias or error.	1. <i>Salience</i> : processes of care may have little meaning to patients unless the link to outcomes can be explained.
2. <i>Easily interpreted</i> : utilization rates of different technologies can often be interpreted by reference to the evidence base rather than necessarily needing inter-unit comparisons.	2. <i>Specificity</i> : care processes are often quite specific to a single disease or single type of medical care so that process measures across several clinical areas or aspects of service delivery may be required to represent quality for a particular group of patients.
3. <i>Smaller sample size</i> : compared to outcome indicators, process indicators can identify significant quality deficiencies with much smaller sample sizes.	3. <i>Ossification</i> : a focus on process may stifle innovation and the development of new modes of care.
4. <i>Unobtrusive</i> : care processes can frequently be assessed unobtrusively (e.g. data stored in administrative or medical records).	4. <i>Obsolescence</i> : the usefulness of process measures may dissipate as technology and modes of care change.
5. <i>Indicators for action</i> : failures identified in the processes of care provide clear guidance on what must be remedied to improve health care quality. They are also more quickly acted upon than outcome indicators, which often only become available after a long time has elapsed.	5. <i>Adverse behaviour</i> : process indicators are relatively easily manipulated and may give rise to gaming and other adverse behaviours.
6. <i>Coverage</i> : process measures can capture aspects of care (such as speed of access and patient experience) that are often valued by patients apart from health outcomes.	

Table derived and expanded from: McGlynn [1998], Davies & Crombie [1999], and Mannion and Davies [2002a].

**Table 2: comparison of the relative advantages and disadvantages of OUTCOME indicators**

<i>Advantages</i>	<i>Disadvantages</i>
1. <i>Focus</i> : a focus on outcomes directs attention towards the patient (rather than the service) and helps nurture a ‘whole system’ perspective.	1. <i>Measurement definition</i> : while some aspects of outcome are relatively easily measured validly and reliably (e.g. death) others are notoriously difficult (e.g. wound infection).
2. <i>Goals</i> : health outcomes more clearly represent the goals of care and the NHS.	2. <i>Attribution</i> : outcomes may be influenced by many factors that are outside the control of a health care organisation.
3. <i>Meaningful</i> : outcomes tend to be more meaningful to some of the potential users of clinical indicators (patients, purchasers).	3. <i>Sample size</i> : outcome assessment requires large sample sizes to detect a statistically significant effect even when there are manifest problems with the processes of care.
4. <i>Innovation</i> : a focus on outcomes means providers are encouraged to experiment with new modes of delivery to improve patient care and experience.	4. <i>Timing</i> : outcomes may take a long period of time to observe.
5. <i>Far sighted</i> : an outcomes focus encourages providers to adopt long-term strategies such as health promotion, which may realise longer-term benefits.	5. <i>Interpretation</i> : observed outcomes may be difficult to interpret if the processes that produced the outcome are complex or occurred distant to the observed outcome.
6. <i>Manipulation</i> : outcomes are less able to be manipulated than process indicators - although providers can influence risk adjusted outcome by exaggerating the severity of patients (upstaging).	6. <i>Ambiguity</i> : good outcomes can often be achieved despite poor processes of care (and indeed <i>vice versa</i> ).

Table derived and expanded from: McGlynn [1998], Davies & Crombie [1999], and Mannion and Davies [2002a].

Clinical indicators of the type described (whether examining processes or outcomes of care) *necessarily* have a comparative element because only rarely can they be assessed in isolation [Davies & Crombie, 1997; Powell et al., 2003]. Such comparisons may be between various provider units (see later), against benchmarks (perhaps derived from evidence-based criteria) or for the same unit across historical time periods. It is this comparative element that provides both the persuasive power of CCIs (their information content) but also poses the most significant methodological challenges (these are explored in detail shortly).

### ***Potential users and potential uses***

Comparative clinical indicators have many potential *users* and can be put to many different *uses*, with users and uses being closely interconnected. Broadly, users can be placed into two distinct categories: clinical practitioners whose *own practice* is being examined by the CCIs; and those other stakeholders (managers; commissioners; regulators; government; the public) who see such data as providing an ‘account’ of clinical practice and performance. Such a distinction is important (and rooted in a principal/agent understanding of the relationship between clinicians and other

stakeholders<sup>1</sup>) because it identifies a wide range of distinctive potential uses of these data. Uses may range from bottom-up, voluntary, professionally-led, quality improvement activities (carried out by the first category of users, clinicians) to top-down, imposed, mandatory, performance management initiatives (driven by various groups in the second category of users). Thus the *same set* of CCIs may, for example:

- act as prompts to reflection on local clinical practice;
- assist in the development of local quality improvement and clinical governance initiatives;
- help to highlight (to clinicians *or* managers/policy people) areas of practice or service delivery in need of further attention;
- be used as part of performance management, regulatory processes or accountability review;
- inform service re-design or policy change;
- contribute to public debates on service quality, performance and accountability.

As we move through this continuum of uses, the key users shift from those most closely concerned with health care practice to those more involved in its management or oversight. At the same time, the degree of engagement shifts from voluntary to more mandatory, as does the role of incentives, moving from intrinsic to extrinsic drivers.<sup>2</sup> At the heart of these shifts is a change in the intent of usage from one focused on *local learning* to one more concerned with *making judgements* (although, of course, such judgements might ultimately be used in various ways to alter service delivery for the enhancement of care).

This identification of various potential *users* and divergent potential *uses* is important because different users/uses place different demands on CCIs in terms of both of our key challenges (robust and interpretable metrics; and systems that encourage data usage). For example, when the focus is on learning we can be less demanding of our metrics' technical properties than when the system leans towards judgement. Similarly, the range of concomitant activities and initiatives associated with the production of CCIs will be radically different if learning is the aim rather than judgement. Other core differences between improvement and accountability approaches to indicator systems are summarised in Table 3 (derived from Freeman [2002]).

---

<sup>1</sup> A principal/agent analysis [Mannion & Davies, 2002b] sees actors in the health care system as either 'principals' (those who are seeking to achieve some goals) or 'agents' (those who are tasked to deliver those goals). The two key concerns that arise in this relationship are (1) 'divergence of objective functions', that is, a recognition that agents might wish to achieve different goals from those specified by the principals, and (2) 'asymmetry of information', that is, acknowledgment the agents usually know much more about the extent to which goals could be met than do principals. Much of the underpinning logic around use of 'performance indicators' *assumes* the first of these concerns is true and is then aimed at addressing the second concern, reducing asymmetries of information. Comparatively less attention has been paid to aligning the goals of principals and agents in the first place.

<sup>2</sup> See later section on *motivation and incentives*.

**Table 3: Differences between accountability and improvement approaches to indicators**

Extended and adapted from Freeman [2002]

	<b>Accountability approaches</b>	<b>Improvement approaches</b>
<b>Emphasis:</b>	Measurement orientated; favouring verification and assurance.	Insight and change oriented; favouring learning to promote continual improvement.
<b>Rationale:</b>	To provide external accountability and ensure/renew legitimacy.	To promote internal change and continuous quality improvement.
<b>Culture:</b>	Comparisons drawn in order to make summative judgements on quality.	Comparisons drawn with a formative emphasis aimed at learning from difference and diversity.
<b>Data presentation:</b>	Data presented as league tables inviting naming, blaming and shaming.	Data presentation emphasises informal benchmarking and acknowledges ‘causal ambiguity’.
<b>Precision required:</b>	High precision needed.	Lower precisions acceptable.
<b>Epistemology:</b>	Empirical. Validity and reliability important alongside statistical assessments of difference.	Interpretive. Use of other data sources and local information acceptable to provide contextual and qualitative understandings.

Given the commissioner for this review (NHS QIS) and the brief (a focus on quality improvement rather than performance management), our emphasis will be on the creation of CCIs that encourage learning – i.e. improvement approaches, rather than accountability approaches. However, in taking this stance we need to acknowledge that CCIs may be taken up and used in ways that are not part of their core purpose or design intent. Media spotlights in particular may fall on any published outputs, and any media reporting is likely to focus on negative rather than positive stories. There is also the potential that, over time, CCIs produced for one purpose may begin to get incorporated elsewhere, for example into more formal process of accountability review, as part of performance management strategies, or perhaps linked to new financial arrangements (the Quality Outcomes Framework, part of the new General Medical Services contract, is just one example of how clinical quality data are becoming more used in performance assessment and reimbursement [Smith, 2004]). Thus it may not be possible for those that design national systems of CCIs to fully control the use to which those data will be put. This review therefore highlights the need for clarity of purpose and ultimately draws attention to the potential dysfunctional consequences that may arise when CCIs garner wider public attention or are adopted for uses beyond their original intent.

## **Making sense of comparative data: creating robust and interpretable metrics**

The logic underpinning the use of CCIs is that we want to be able to assert that variation across indicators reflects something real and important about the underlying delivery of health care. That is, we wish to assert, or at least to suggest as a working hypothesis, that those whose measures fall at the ‘less desirable’ end of the scale may, at the very least, need to look further at the reasons for this. However carefully such sentiments are expressed, they all embody the notion that there is some meaningful degree of causality between the services delivered and the ‘scores’ obtained on any CCIs. The way in which the indicators are developed, defined, operationalised and analysed will have a big impact on the credibility of this logic. This section explores many of the issues arising in the development and analysis of CCIs that make them more or less interpretable. Given limitations on space, the nature of each issue is described relatively briefly (see cited references for more detailed explication and empirical exemplification, although it should be noted that these are illustrative rather than comprehensive).

### ***Metric selection and design***

There is growing attention now being given to more systematic development of quality indicators, particularly when these are based on process measures [Campbell et al., 2003; Mainz, 2003]. When the evidence base is strong about key clinical processes then many obvious candidates for CCIs present themselves. When, as is more usually the case, evidence is weaker, systematic methods for combining evidence and expert opinion in the development of appropriate CCIs have been developed [Campbell et al., 2003; Clarke & Rao 2004]. These include consensus development conferences, Delphi methods, the nominal group technique, and the RAND appropriateness method (see Campbell et al. [2003] for more details and further references explaining these techniques). These approaches can help address both face validity (professional consensus) and content validity (by explicit inclusion of available evidence). As such they may help in the creation and specification of CCIs that garner professional interest.

Rather unjustly neglected in many current discussions is the extent to which any set of CCIs also reflects the core priorities and concerns of patients and users [Davies & Marshall, 1999]. Thus lay input and patient/user involvement in the design and selection of CCIs may also be important but is rarely systematically addressed. In addition, since development of CCIs has been an international effort, there is considerable scope for adopting and adapting indicators that have been developed elsewhere [Steel et al., 2004; Clarke & Rao 2004] – useful indicators may not always need to be developed from scratch.

### ***Measurement properties***

Any measure selected must pass muster against a wide range of technical criteria [Powell et al., 2003; Campbell et al., 2002]. Key amongst these are *validity* (does the measure capture what it is supposed to?), *reliability* (how reproducible are the measures when taken at different times or in different circumstances?), and *sensitivity to change* (will the measure discriminate between good and poor quality care and be able to detect perhaps small but worthwhile improvements?). None of these criteria admit to simple yes/no answers, and the extent to which any measure meets them can be difficult to assess and open to considerable levels of disagreement. These problems are most

significant when judgements have to be made on what is being observed; for example, an analysis of wound infection rates found three-fold variations between rates when different definitions were used. It concluded that even with a fixed definition ‘*differences in interpretation prevent comparison between different centres*’ [Wilson et al., 2004]. Even apparently ‘hard’ endpoints such as death can be problematic [Beaumont & Hurwitz, 2003], with changing definitions over time periods and sources of data collection having significant impacts on apparent mortality rates [Thomas & Hofer, 1999; Goldacre et al., 2002; Mohammed et al., 2004].

Deficits on these technical criteria may be significant and systematic, for example, when data are collected unblinded there is evidence that this affects the judgements that are reached in systematic ways [Noseworthy et al., 1994]. Indeed, any changes to the detail of by who and how data are collected at local level may influence the extent to which the measures are valid, reliable and sensitive. These desirable technical characteristics of indicators can be broadened to cover other important issues such as *acceptability* (helped by the consensus methods to indicator development outlined above), *feasibility* (an acknowledgement of the constraints placed by data availability) and *communicability* (the degree to which the meaning of the indicator can readily be explained). Indeed, one research team suggested that up to 12 attributes for quality indicators ought to be considered [Pringle et al 2002; see Box 1 below]

**Box 1: Desirable attributes for quality measures**

Adapted from Pringle et al [2002]

- Face validity:* Key stakeholders must see that doing well on the quality measure represents better quality care;
- Communicable:* Relevance of measure can be easily explained and understood by key target audiences;
- Effective:* Measure measures what it purports to measure, and is free of perverse incentives;
- Reliable:* Data underpinning measure should be complete, accurate, consistent and reproducible;
- Objective:* Data should be as independent of subjective judgement as is possible;
- Available:* Data should be collected for routine clinical or organisational reasons and be available quickly with minimum extra effort or cost;
- Contextual:* Measure should be context-free, or important context effects should be adjusted for;
- Attributable:* Measures should reflect the quality of care by individuals, teams and organisations;
- Interpretable:* Measures should be capable of ready interpretation of core underlying system performance;
- Comparable:* Measures should allow reliable comparison with external benchmarks or to other data sets gathered in similar circumstances;
- Remediable:* Measures should point to actionable areas for improvement that are likely to bring about changes in the measures recorded;
- Repeatable:* Measures should be sensitive to improvements over time.

### ***Unit of analysis***

An issue that has received surprisingly little empirical examination in the literature is the question of what is the appropriate unit of analysis for any given CCI. Several choices present: data can be aggregated at the level of the individual practitioner (e.g. mortality rates for specific surgeons), a clinical team, an organisational unit (service or whole hospital), or a region (geographical area of even whole nation). The benefit of aggregation is that it increases numbers (reducing statistical uncertainty). Perhaps also helpful, it reduces the extent to which individual practitioners can be identified. The downsides are that larger units of analysis may include a greater degree of heterogeneity of practice, make the data less interpretable and perhaps less salient, and therefore obscure the types of changes that such data ought to promote.

A further complication is that more care, especially of chronic disease, is now being delivered through networks of professionals that cut across traditional organisational boundaries [Thomas, 2003; Christie, 2003]. If these new models of health care delivery are to be able to access the benefits of CCIs then new units of analysis will be needed which focus on catchment populations rather than organisations *per se*. Since data gathering and sharing are key concerns of managed clinical networks (MCNs) it should be possible to achieve this but how this dovetails with national data acquisition may need careful exploration.

### ***Case-mix, confounding and case-mix adjustment***

Comparisons across quality indicators need to take into account the degree to which different ‘units of assessment’ (whether individuals, teams or organisations) are dealing with similar patient groups. If this is not done, differences seen across indicators may be more attributable to the sorts of patients being dealt with (case-mix) rather than any divergence in practice. This is especially important when examining health outcomes [Davies & Crombie, 1997; Powell et al., 2003], but even process measures may vary as a function of the patients seen [Crombie & Davies, 1998]. A wide range of patient characteristics may need to be taken into account. At the most basic level, adjustments for age, sex, disease severity and co-morbidities are all important, but much more sophisticated systems of case-mix adjustment have been developed, some of which use dozens of variables [Iezzoni et al., 1994; 1997; Kuhlthau et al., 2004]. Developing and testing such statistical adjustment schemes is highly technical and complex, and poses additional data demands (with familiar concerns for the validity and reliability of those measures which form the basis of any adjustments).

Case-mix adjustment methods are necessarily incomplete – even the best models linking patient characteristics to patient outcomes are able to explain only a portion of the variation. Research has also shown that different adjustment schemes may lead to different rankings across key indicators [Iezzoni et al., 1994; 1997]. Thus it can be unclear in advance which is the ‘best’ adjustment method to use, and different methods may lead to different conclusions about the quality of care in individual units.

Case-mix adjustment is not only a complex technical problem in itself – it is also a dynamic concern. Once attention is focused on those aspects of case mix that are being adjusted, there is a danger that these will be assessed and/or recorded differently. This process, called ‘upstaging’,

category ‘drift’ or diagnosis ‘creep’, often sees a gradual shift towards the grading of patients as more severe together with a greater degree of diligence in the identification of co-morbidities [Powell et al., 2003; Green & Wintfeld, 1995]. If this process occurs, we can have less confidence in any comparison, either historically in the same unit or indeed cross-sectionally across units (where the process may operate differentially).

### ***Coping with chance variability***

Chance variability is present in all quantitative data and CCIs are no exception. Traditional statistical techniques such as assigning p-values (hypothesis testing) or adding confidence intervals to estimates can aid interpretation amongst the statistically literate, but are confusing to many stakeholders and are themselves open to pitfalls. In interpreting CCI data we can be misled in two main ways. First, our attention can be drawn to differences that do not in fact represent real underlying differences but are instead the result of random variation. These (Type I) errors become more likely the more statistical comparisons are made (unless special techniques are engaged to adjust p-values or confidence limits [Seneta & Chen, 2005]). This can be a particular problem with CCIs, which are often many and invite multiple comparisons across many units of assessment. While falsely identifying any particular provider as unusually able may not cause undue concern, the obverse, falsely denigrating providers as ‘poor performers’ may have considerable deleterious effects [Lilford et al., 2004]. Such false denigration may not in fact be rare: careful simulation work examining mortality rates after heart attack [Hofer & Hayward, 1996] suggested that up to three-quarters of providers labelled as ‘poorly performing’ may in fact be no worse than average; other work examining the predictive power of early readmission rates in cardiac disease found around two-thirds of hospitals were falsely identified as poor [Hofer & Hayward, 1995].

Second, we may be misled into believing that no differences exist between units when measures all fall within statistical limits. If the number of events being observed is relatively small we may be being falsely reassured: real and important differences may lie undiscovered, obscured by random fluctuations (a Type II error). In practice, it is not possible to reduce simultaneously the chances of both errors unless we increase the overall number of events being observed. Since this may mean that we then have to enlarge our units of assessment or aggregate across additional years of data, reducing the play of chance is often achieved only at the expense of reducing the focus (and hence the salience) of the analysis.

Assessing changes over time in CCIs can also be difficult and is prone to the same two errors (Type I and Type II) outlined above. When data are presented in rankings, large shifts in rank over time (either up or down) may not in fact represent statistically significant changes in the underlying measure [Marshall & Spiegelhalter, 1998]. Moreover, the statistical phenomenon of ‘regression to the mean’ [Bland & Altman, 1994] will also account for a portion of any apparent movement back to average from outlier status. While technical methods are available to assess the extent of this [Hayes, 1988], they are not easy to apply and have rarely been used with CCIs.

Other more sophisticated techniques have been developed to address chance variability, such as statistical process control charts (to disentangle common cause from special cause variation [Mohammed et al., 2001; Morton, 2003]) and multi-level modelling (to separate out hierarchical effects [Goldstein & Spiegelhalter, 1996; Spiegelhalter & Leyland, 2001]). These seem to offer

potentially fruitful avenues, but despite being promoted over a number of years [Thomson & Lally, 2000; Mohammed et al., 2001; Morton 2003] they have yet to see widespread application to routine data in health care.

### ***Data quality and completeness***

It is self-evident that even the most well-chosen, carefully operationalised, and rigorously risk-adjusted indicator will be undermined if the data sets on which it draws are suspect. Yet research has shown repeatedly that routine data sets are frequently incomplete and inaccurate [McCarthy et al., 2000; Naessens & Huschka, 2004]. While such deficiencies can be expected to improve with increased investment and a greater realisation of the importance of such data, they are unlikely ever to be fully eradicated. Moreover, even as completeness and accuracy are increased, local variations in precisely how data are gathered, collated and analysed are likely to persist and these may have subtle and enduring effects. Such data deficiencies are problematic because, wherever there are questions over data quality and completeness, experience shows that this is the first area to be explored by those seeking to explain away any apparent variation in CCIs [Davies, 2001; Narayn-Lee et al., 2004; Mannion et al., 2005].

A further issue that may undermine confidence is when centrally collated and analysed data conflict with the local data systems. Research has shown that CCIs created from different data sources (e.g. data gathered from clinical versus administrative records) often tell quite divergent stories about the quality of local care services [McCarthy et al., 2000; Scott et al., 2004; Mohammed et al., 2004; Mukamel et al., 2004; Spies et al., 2004]. An important aspect of gaining acceptance therefore will be the need to demonstrate some degree of convergence between local and national data sets or, failing this, plausible explanations for any divergences seen.

### ***Concluding comments on developing interpretable metrics***

In sum, there are many issues that need to be addressed in developing credible metrics with a high degree of interpretability. The observational nature of the data, and formidable technical problems, place many obstacles in the way of interpretation before variation across CCIs can be construed as reflecting real differences in the quality of care (See Box 2 below, derived from Powell et al [2003]).

## Box 2: Variations in measured quality: real difference or artefact?

Derived from Powell, Davies & Thomson [2003]

When CCIs created from routine data reveal variations between different service providers in reported quality measures, this *may be* evidence of real differences in quality of care.

However, other possible causes of variation include:

- **Problems with measurement** - validity and reliability of measures are undermined by:
  - inappropriate/insensitive data sources – e.g. data taken from administrative systems may lack necessary clinical details
  - measures which are too narrow to reflect the breadth of care provided – e.g. using a single measure in psychiatric care
  - inappropriate/insensitive definition of outcomes – e.g. looking at 30 day mortality for a particular condition even when most deaths fall outside this period
  - changes in data recording over time – e.g. apparent improvement or deterioration because of changes in reporting practices
  - differences in data recording between providers – e.g. data collection processes may not be equivalent and may lead to apparent variations
  - lack of blinding – e.g. unblinded assessment of outcome is prone to bias
- **The presence of case mix and other factors** – apparent differences between units may be more attributable to differences in the patient groups (e.g. in clinical and socio-demographic terms, and in terms of contextual factors) than to any true differences in performance. Yet case-mix adjustment is demanding:
  - it is always incomplete as adjustment can only be made when the necessary data are available and when the relevant factors are known
  - the choice of adjustment scheme can itself affect quality rankings
  - all adjustment schemes risk ‘upstaging’ where the severity grading of patients may drift upwards over time, with implications for severity-adjusted outcomes
- **Chance variability** – this can lead to falsely identifying outliers for praise/blame (‘Type I errors’) or can obscure real differences and hide poor performers (‘Type II errors’)
- **Poor data quality** – despite growing awareness of the problem, routine data systems are often incomplete or inaccurate, which can undermine any conclusions drawn

These difficulties are well illustrated in a recent analysis of historical data for a cohort of 143 surgeons that included an individual gynaecologist (Rodney Ledward) who was subsequently confirmed to have been performing very significantly below what might reasonably have been expected [Harley et al., 2005]. Seven clinically relevant indicators were chosen from routinely available data sets, and complex multi-variate analysis (calculation of what is termed the Mahalanobis Distance) showed that this individual practitioner was indeed an outlier in these data sets *but only in three of the five years studied*. In addition, the proportion of other surgeons who were identified as statistical outliers in any one year ranged from 9-20%. Without additional information, these data sets illustrate both false reassurance (those years when a known poor performer was *not* a statistical outlier) and a large number of potential false denigrations (unless one is to believe that up to 20% of the surgeons in this cohort are indeed ‘poor performers’).

Notwithstanding these difficulties, which limit the degree of certainty we may have over our readings of the data, when CCIs are used in systems of learning rather than judgement a high degree of ‘causal ambiguity’ can be acknowledged and accepted without precluding further investigation and quality improvement activities. It is to this issue that we next turn.

## **Encouraging effective data usage**

This section explores the design issues arising when considering how best to share CCIs. It addresses crucial questions of who should be the key target audiences for these indicators, how should the data be presented, and what contextual factors need to be considered (such as incentives for metric use and availability of additional analytic support). Finally, consideration is given to the potential for dysfunctional responses to data publication, and how these unwanted responses might be mitigated.

### ***Evidence and audience: who uses comparative clinical indicators?***

The evidence to-date on the uptake and use of CCI data is not very encouraging [Mannion & Goddard, 2001; 2003; Lanier et al., 2003; Sheldon 2005]. In general, centralised systems of analysis and reporting of CCIs have been shown to have only minor effects on quality of care, although this may reflect more an absence of evidence than evidence of absence [Marshall et al., 2000; Marshall et al., 2003; Sheaff et al., 2003; Jamtvedt et al., 2005]. This over-arching finding is however consistent with Cochrane reviews of the effects of audit and data feedback more generally: these approaches are sometimes seen to improve performance, but any effects are variable, and usually small to moderate [Lanier et al., 2003; Jamtvedt et al., 2005].

The public generally, and ‘consumers’ of health care more specifically, while expressing interest in these data, do not appear to use them to any great extent when making decisions [Marshall & Davies, 2001; Schneider & Lieberman, 2001; Marshall et al., 2002; Mannion & Goddard, 2003]. Indeed one review of the evidence specifically ruled out the mechanism of ‘consumer choice’ as the means through which provider performance is improved [Sheaff et al., 2003]. Furthermore, referring physicians (including general practitioners) also seem to make little use of the data [Mannion & Goddard, 2004], even when these data seem to offer very specific guidance (for example, referring cardiologists have not been shown to use the data that identify surgeons with high and low mortality rates, even when referring patients for specific operative procedures [Schneider & Epstein, 1998]). Health care purchasers and commissioners have begun to show increasing interest in these data (especially in the United States, where major employers are seeking to use the data to secure better care for their enrolled employees) but more often purchasing decisions are driven by cost and local availability of service [Marshall et al., 2003]. The two areas that seem to offer the greatest potential for use are first, application by individual physicians whose work is covered by the data, and second, action by the organisations within which these physicians work.

Physicians are sometimes aware of CCI data when these relate to their own practice, but they are often sceptical of the messages communicated [Davies, 2001; Mannion & Goddard, 2003]. If they are aware of the data (and oftentimes these data do not percolate to service levels unless they indicate the presence of a serious problem [Mannion & Goddard, 2001]) initial responses can be

defensive, involving attempts to ‘explain away’ differences using some of the issues outlined above: validity, reliability, case-mix, chance variability and, particularly, data quality [Mannion & Goddard, 2003]. There is also evidence that physicians’ strongly held prior beliefs may change very little, even when faced with detailed and credible data [Parry & Tucker, 2004]. Thus we should remain modest in our expectations of the degree of change that will be prompted solely by the sharing of indicator data [Jamtvedt et al., 2005].

In contrast to the relatively muted responses from individuals, the overarching organisations within which physicians are located (which may or may not have an employing relationship with those physicians) can often prove the most sensitive of recipients for CCI data [Marshall et al., 2000; Davies, 2001; Mannion & Goddard, 2003]. They may initially seek to discredit the data, but can then move on to make more positive responses, for example, improving internal data systems, encouraging quality improvement activities, or reassigning clinical staff. This is encouraging since a systems view of quality and safety in tandem with the recognition of the need for cultural changes [Davies et al., 2000] suggests that it is provider organisations who hold the key to quality improvements in health care. Thus the findings that health care organisations can be responsive to clinical indicator data suggest that these organisations together with their front-line clinical staff should be the prime targets for CCI data.

### ***The importance of credibility***

Qualitative studies of how clinicians respond to data suggest the importance of having credible quality indicators with ‘scientific respectability’ [McGlynn, 1998] in tandem with an honest portrayal of data limitations [Davies, 2001; Mannion & Goddard, 2003]. Such an approach emphasises an openness to various and diverse ‘readings’ of the data, and an unwillingness to draw specific inferences too early in the process of interpretation. These in turn imply: careful metric selection (data that relate to national goals and priorities but also speak to front-line clinical concerns); credible, accurate and up-to-date data (lags of up to three years are not uncommon [Broder et al., 2004]); metrics properly adjusted to give value-added information; and data presented non-judgementally for local interpretation.

Additional important features here may be the means by which the data are displayed and communicated, and the extent to which the data suggest local courses of action. The ways in which the data are presented may influence the type and degree of user engagement [Gysels et al., 2004]; for example, data presented as control charts, when compared with league tables, prompted fewer requests for additional contextual information and led to fewer providers being labelled as outliers [Marshall et al., 2004]. Similarly, data presented in ways that illustrate and quantify the potential gains possible from quality improvements may also do more to engage front-line stakeholders [Gibberd et al., 2004]. In addition, data that appear to front-line clinicians as readily interpretable and, as important, data that suggest courses of action that are likely to yield service improvements are more likely to be shared, discussed and acted upon [Marshall et al., 2004].

Finally, producers of centralised systems of data analysis need to recognise that other data sources, especially local data systems, may not always tell stories congruent with national data sets. When national and local data sets do triangulate this may provide an important stimulus to action. Thus

centrally produced CCIs need careful and sensitive interpretation that integrates local perceptions driven by local/regional data sets and day-to-day experience.

### ***Paying attention to motivation and incentives***

If sharing CCIs is to bring about improvements in health care then the patterns of behaviour of targeted audiences must be changed. This in turn highlights the importance of both *motivation* (the internalised desire for action) and how this might be elicited i.e. the role of *incentives*. Both of these concepts (motivation, incentives) are complex, and the relationships between them, and between each and subsequent patterns of behaviour, are similarly difficult to disentangle [Smith, 2004; Marshall & Harrison, 2005].

Core drivers to motivation are often categorised as either intrinsic or extrinsic. *Intrinsic* drivers are those internalised aspects of values and cognition that motivate particular patterns of behaviour, for example, moral choices about ‘doing the right thing’, pride in ‘a job well done’, or meeting the expectations of peers. In comparison, *extrinsic* drivers are those external factors that shape behaviour such as financial rewards, regulatory demands or fear of reputation damage. The evidence is pretty clear that extrinsic drivers – such as financial incentives, or fear of ‘naming and shaming’ – can elicit substantial shifts in behaviour patterns [Smith, 2004; Marshall & Harrison, 2005; Conrad & Christiansen, 2004]. What is less clear is the extent to which an emphasis on the extrinsic may deplete or damage intrinsic drivers [Marshall & Harrison, 2005]. For example, much of health care has typically been delivered by health care professionals on a *discretionary basis*; however as relationships between these professionals and their supporting organisations become more *contractual*, this begs important questions about whether extrinsic drivers will be responsible for ‘crowding out’ important intrinsic drivers.

Work in the US [Rosenthal et al., 2004; Conrad & Christiansen, 2004] has highlighted the diverse range of incentives that might be employed to drive quality improvement work, identifying different types of both financial and non-financial levers. Financial incentives might include, for example, quality bonuses, performance-related fees, or compensation withholds, as well as quality grants, access to performance enhancement funds or reimbursement for care planning, shared care or other activities that are likely to enhance quality. Non-financial incentives include performance profiling with or without media publicity, technical assistance for quality improvement, or various types of ‘earned autonomy’ such as reductions in administrative or regulatory requirements.

This focus on incentives does tend to privilege a conceptualisation of health care professionals as ‘rational calculating individuals’. This stance has in fact been much criticised as incomplete or even misleading, with invitations to consider the importance of value-led behaviour [Davies & Lampel, 1998; Davies et al., 2000]. If this is done, the core problem becomes less one of reducing information asymmetries and incentivising, and more one of understanding and aligning the values, beliefs and assumptions of various stakeholder groups; in short, a consideration of the sorts of cultural changes that would be needed to bring about assiduous and joined-up efforts at quality improvement in health care organisations [Ferlie & Shortell, 2001; Davies et al., 2000].

While some of this may seem a little far removed from the production and sharing of CCIs, the potential for ‘mission drift’ in the use of CCIs discussed earlier, and the ever-present potential for

media spotlighting, urge some degree of consideration of inadvertent effects. In addition, there is some need for the designers of systems of CCI dissemination to pay attention to the incentive frameworks within which these indicators are to be distributed if they are not to be (as often happens) largely ignored. Thus there need to be sufficient incentives around the CCIs to encourage motivation to pay attention, take seriously, and act upon data; but drivers not so strong that they upset important components of intrinsic motivation, or induce some of the potential dysfunctional components reviewed later.

### ***Support with tools for further analysis and action***

Possibly the greatest consideration in any national system of CCI analysis and dissemination is how to ensure that the measures created are picked up and used in those environments where change is needed to bring about service improvement. The evidence to date (outlined above) suggests that even when attention has been given to all of the issues so far identified there is every possibility that the data will go largely unconsidered at the front line. Given this, it would seem imperative that a first step should be systematic enquiry into the information needs, data presentation preferences, and associated analytic support, that might be required by different audiences. In making these enquiries, potential audiences should be differentiated in at least four ways: their position and role in the provider organisation (broadly, clinician or manager); their profession (doctors, nurses and allied health professionals are likely to have different skills and information needs [Harvey, 2004]); their clinical specialty (contexts vary widely between, for example, intensive care, psychiatric care and primary care); and their degree of skills, aptitude and enthusiasm (programmes designed only to appeal to ‘early adopters’ will struggle to engage the wider majority).

Building on a clear understanding of audience needs, any national programme sharing CCIs should be very clear about how the data are meant to be disseminated and cascaded within the service. Evidence over many years suggests that passive dissemination (public release, combined with mail or electronic circulation) are unlikely to be very effective [Jamtvedt et al., 2005]. More likely to have effect are active programmes of diffusion that engage key stakeholders in face-to-face discussions with the data, alongside a clear articulation of *how* the data might be used by those stakeholders. Thus, the organisations who are recipients of CCI data should know *who* these data should be shared with, and *how* they might be actioned and supported locally.

Providing the means to support data usage at local level will be a significant challenge. Some possibilities to encourage more active engagement include:

- A programme of workshops and seminars where the data are presented and opened out for discussion with clinicians and other stakeholders;
- The commissioning of expert analysis and interpretation that sits alongside the data and explores possible meanings and courses of action (being mindful that such analysis should *open up* rather than *close down* interpretive avenues);
- The development of web-based facilities that allow access to the raw data and analytic tools to allow further user-driven analysis and exploration;
- Developing partnerships with funding bodies and change agencies so that local clinicians wishing to explore and action findings from CCIs have ready routes into supportive funding and relevant expertise;

- Strengthening partnerships between different programmes in NHS QIS to enable better linkages between the evidence base underpinning high quality care and the data sets that document the care that is delivered (CCIs). In this way local providers can be encouraged to see NHS QIS activities as seamless, mutually supportive and synergistic.

Whether sharing of CCIs will be effective in driving quality improvement is critically dependent on the organisational contexts where such measures are distributed [Bradley et al., 2004]. Thus, in tackling the above agenda, it will be necessary to know much about current local arrangements for quality improvement and clinical governance in the targeted organisations. The challenge will be to aim for complementarities and synergies rather than, as is often the risk, disconnected initiatives that can duplicate and sometimes conflict with existing arrangements [Ferlie & Shortell, 2001]. As a result, engagement with provider organisations may need to be both intensive and bespoke. As local arrangements differ widely, engagement strategies should reflect this. One size is unlikely to fit all and, moreover, such variation of local contexts is as likely to be seen within organisations as between them [Renshaw & Ireland, 2003].

Bringing about change in organisations is far from simple and any discussion here must acknowledge the pervasive issues of organisational power and politics, in particular professional power (especially that of the medical profession) and how this relates to managerial power [Davies & Harrison, 2003; Davies et al., 2003; Gray & Harrison, 2004]. Yet an extensive literature now provides helpful insights into various facets of change and improvement in health care organisations: models of change and change management have been reviewed and summarised [Iles & Sutherland, 2000], the literature on how innovations are adopted and spread within organisations offers further insights [Greenhalgh et al., 2004], as does a recent review on theories of behaviour change that can be used to guide quality improvement [Wensing et al., 2005]. Other work has more specifically addressed the needs of clinical governance [Degeling et al., 2004], the evidence base supporting quality improvement efforts [Shojania & Grimshaw, 2005], and the organisational change theory underpinning use of indicators (in this case, in general practice [Rhydderch et al., 2004]). Given the wealth of this knowledge and its disparate sourcing this is a valuable, but nonetheless difficult, resource that can be tapped. As NHS QIS plans for supporting users are clarified it is suggested that further review work might help to ensure that any package of measures is properly underpinned by theory and evidence of service-level change.

### ***Avoiding dysfunctional consequences***

All policy and managerial initiatives have the potential for unwanted, unexpected and dysfunctional consequences. Those arising from the publication and dissemination of performance-related data have been relatively well explored in the literature [Smith 1995; Smith 2002; Goddard et al., 2000]. Box 3 lays out a range of unintended responses to the greater availability of CCIs. These range from the undesirable but not actively destructive (myopia and convergence) to the out-right corrupting (misrepresentation and fraud). The ways in which such data are shared, and the implicit assumptions (about judgements) embedded in the systems of communication and support, can be very influential in terms of promoting or damping down the potential for such dysfunctional consequences (*'systems that do not trust people beget people that cannot be trusted'* [Davies & Lampel, 1998]).

### **Box 3: Possible dysfunctional consequences**

Adapted from Smith [1995] and Marshall et al. [2004]

Organisations or individuals may alter their behaviour in a variety of undesirable ways:

- They may concentrate on the clinical areas being measured to the detriment of other important areas ('tunnel vision');
- They may pursue narrow organisational objectives at the expense of strategic coordination ('sub-optimisation');
- They may concentrate on short-term issues and neglect long-term criteria ('myopia');
- They may place greater emphasis on not being exposed as an outlier rather than on a desire to be outstanding ('convergence');
- They may be disinclined to experiment with new and innovative approaches for fear of appearing to perform poorly ('ossification');
- They may alter their behaviour to gain strategic advantage ('gaming');
- They indulge in selective and creative data gathering, classification and coding, perhaps extending to outright misreporting and fraud ('misrepresentation').

While it can be hard to assess the extent to which dysfunctional consequences are likely, and difficult to evaluate the extent to which they are occurring in practice, their presence or otherwise is something that needs regular attention and review. Box 4 identifies a number of ways in which the potential for dysfunctional consequences may be reduced, but it is readily seen that there are certain tensions here; for example, keeping the number of indicators small and manageable is in direct conflict with the need to have a balanced portfolio of indicators that represents the interests of users and carers as well as health care professionals. Nonetheless, the guidance presented here does provide the basis for some discussions as to how this difficult and troubling aspect of data sharing might begin to be addressed. Crucially, any dysfunctional responses are likely to become more manifest over time as those whose practice is being described become more knowledgeable and adept at deflecting or neutralising any critical potential [Smith, 1995]. Thus any scheme disseminating CCIs will need regular review to assess the extent to which any benefits are being attenuated in favour of dysfunctional consequences. As the recent Working Party on Performance Monitoring in Public Services commented: *'performance monitoring done well is broadly productive for those concerned; done badly, it can be very costly, and not merely ineffective but harmful and indeed destructive'* [Bird et al., 2005].

### **Box 4: Ways of reducing potential dysfunctional consequences**

Adapted from Smith [1995] and Marshall et al. [2004]

- Ensure that staff are involved at all levels in the organisation;
- Be flexible in how the measures are used;
- Keep the number of indicators small and manageable;
- Ensure that the portfolio of indicators includes a balanced selection covering outcomes and client satisfaction as well as those indicators of core professional interest;
- Make use of independent benchmarks if possible (e.g. evidence-based criteria);
- Seek expert interpretation of the indicators, encompassing both local and other external and independent experts;

- Keep the reporting systems under constant review;
- Take a longer-term perspective, with a rolling programme that seeks progressively more inclusive buy-in;
- Highlight the importance of continuous learning over snap-shot judgements.

## Concluding remarks

The absence of data on individual professional practice has been identified as a significant block to achieving greater physician engagement in quality improvement activities [Mainz, 2004; Audet et al., 2005]. At the same time, physicians have significant anxieties about the public sharing of individualised data on clinical performance [Audet et al., 2005]. Thus the creation of national systems of CCIs is both a significant opportunity to address an important barrier to quality improvement, and a challenge to do so in a way that does not antagonise, demoralise or induce unintended consequences. This Discussion Paper has laid out a series of issues that require detailed consideration if the potential for benefits is to be maximised and pitfalls are to be avoided. These are summarised in Box 5.

At the heart of these arguments is the need to emphasise critical engagement with those whose practice is being measured, and to reassure them that the data gathered and analysed will be interpreted in the context of learning rather than judgement. Creators of CCIs need to recognise that the approach is inherently backward looking – ‘*like trying to drive a car by looking through the rear-view mirror*’ [quoted in Wilcock & Thomson, 2000] – and may encourage an unnecessarily defensive posture on past practice. The challenge then is to encourage a refocus away from explaining previous performance to one that is forwards looking, learning from the past primarily to improve future care. This can best be achieved through proper engagement with the grassroots and support for local systems of data interpretation and actioning.

### **Box 5: Developers of national systems of CCIs need to think through:**

- The nature of the primary intended audience for CCIs
- How this audience can be engaged in metric selection, definition and interpretation
- How best to generate metrics that meet a wide range of the technical criteria outlined above
- The extent to which any differences are interpretable – and the clear communication of this
- The ways in which the data should be presented, graphically and numerically, in ways that are most communicative and acknowledge chance variability
- What incentives for health care providers (as individuals and as organisations) are sufficient to garner attention, but unlikely to cause too many dysfunctional consequences
- The extent to which nationally produced CCIs can be integrated with local clinician-led data systems
- The potential for active training and support of the media in data interpretation
- The means by which the information contained in CCIs can be actioned by individual professionals and within health care organisations
- How such learning/change can be supported at individual, team and organisational levels
- How dysfunctional responses can be identified and minimised

- How gains can be sustained in a rolling programme that balances continuity of some measures with the introduction of new insights and challenges

It is also worth noting that any national system of CCIs will not exist in a data vacuum. Many other systems of data gathering, analysis and interpretation have grown up aimed at understanding and supporting improvements in practice. These include the National Audits initially funded by the Clinical Resource and Audit Group (CRAG), data gathering connected with the creation and operation of MCNs [Christie, 2003; Thomas, 2003], confidential systems of audit such as the critical event analysis of surgical deaths [Baxter 2005; Thompson & Stonebridge, 2005] and other specialist clinical data systems [Black 1999; Black & Payne, 2003]. As CCIs are developed, it makes good sense to explore thoroughly the availability of other datasets in any given clinical area; examining their contents, distribution, and patterns of use. In this way, national systems might better dovetail with bottom-up systems that have already secured significant local buy-in.

Finally, because of the (to date) only limited evidence on the effectiveness of national systems of CCIs, any new system put in place should also seek to contribute to an evidence base of its own effects. That is, as new CCIs are created and rolled out there is a need for a concomitant programme of research that explores their impacts in their widest sense, including dysfunctional consequences. Properly planned and conducted, such a parallel programme of research could make vital contributions to our understanding about the effectiveness of such schemes. More importantly, it would have the potential to contribute significant on-going learning enabling smarter systems of CCIs and their supporting activities to be developed.

*END*

## References

- Audet, A.-M. J., M. M. Doty, et al. (2005). "Measure, learn, improve: physicians involvement in quality improvement." Health Affairs **24**(3): 843-53.
- Baxter, N. N. (2005). "Monitoring surgical mortality." BMJ **330**: 1098-1099.
- Beaumont, B. and B. Hurwitz (2003). "Is it possible and worth keeping track of deaths within general practice?" Quality & Safety in Health Care **12**: 337-42.
- Bird, S. M., S. D. Cox, et al. (2005). "Performance indicators: good, bad and ugly." Journal of the Royal Statistical Society: Series A **168**(1): 1-27.
- Black, N. (1999). "High-quality clinical databases: breaking down barriers." Lancet **353**(9160): 1205-1206.
- Black, N. and M. Payne (2003). "Directory of clinical databases: improving and promoting their use." Quality & Safety in Health Care **12**: 348-52.
- Bland, J. M. and D. G. Altman (1994). "Some examples of regression towards the mean." British Medical Journal **309**: 780.
- Bradley, E. H., E. S. Holmboe, et al. (2004). "Data feedback efforts in quality improvement: lessons learned from US hospitals." Quality & Safety in Health Care **13**: 26-31.
- Broder, M. S., L. Payne Simon, et al. (2004). "Surgical quality: review of Californian measures." BMJ **328**: 152-153.
- Campbell, S., J. Braspenning, et al. (2003). "Research methods used in developing and applying quality indicators in primary health care." BMJ **326**: 816-819.
- Campbell, S., M. Hann, et al. (2002). "Quality assessment for three common conditions in primary care: validity and reliability of review criteria developed by expert panels for angina, asthma and type 2 diabetes." Quality & Safety in Health Care **11**: 125-130.
- Christie, B. (2003). "Scotland's Health and Health Services." BMJ **327**: 1233.
- Clarke, A. and M. Rao (2004). "Developing quality indicators to assess quality of care." Quality & Safety in Health Care **13**: 248-249.
- Conrad, D. A. and J. B. Christianson (2004). "Penetrating the "black box": financial incentives for enhancing the quality of physician services." Med Care Res & Rev **61**(3): 37S-68S.
- Crombie, I. K. and H. T. Davies (1998). "Beyond health outcomes: the advantages of measuring process." Journal of Evaluation in Clinical Practice **4**(1): 31-38.
- Davies, H. T. (2001). "Public release of performance data and quality improvement: internal responses to external data by US health care providers." Quality in Health Care **10**: 104-110.
- Davies, H. T. and I. K. Crombie (1995). "Assessing the quality of care." BMJ **311**: 766.
- Davies, H. T. and I. K. Crombie (1997). "Interpreting health outcomes." Journal of Evaluation in Clinical Practice **3**(3): 187-199.
- Davies, H. T. and I. K. Crombie (1999). "Outcomes from observational studies: understanding causal ambiguity." Drug Information Journal **33**(1): 153-158.
- Davies, H. T. and M. N. Marshall (1999). "Public disclosure of performance data: does the public get what the public wants?" Lancet **353**: 1639-1640.
- Davies, H. T. O. and S. Harrison (2003). "Trends in doctor-manager relationships." BMJ **326**: 646-9.
- Davies, H. T. O., C. L. Hodges, et al. (2003). "Consensus and contention: doctors' and managers' perceptions of the doctor-manager relationship." Br J Hlth Care Man **9**(6): 202-8.
- Davies, H. T. O. and J. Lampel (1998). "Trust in Performance Indicators?" Quality & Safety in Health Care **7**(3): 159-62.

- Davies, H. T. O., S. M. Nutley, et al. (2000). "Organisational culture and quality of health care." Quality & Safety in Health Care **9**: 111-19.
- Degeling, P. J., S. Maxwell, et al. (2004). "Making clinical governance work." BMJ **329**: 679-82.
- Ferlie, E. B. and S. M. Shortell (2001). "Improving the quality of health care in the United Kingdom and the United States: a framework for change." Milbank Quarterly **79**: 281-315.
- Freeman, T. (2002). "Using performance indicators to improve health care quality in the public sector: a review of the literature." Health Services Management Research **15**: 126-137.
- Gibberd, R., S. Hancock, et al. (2004). "Using indicators to quantify the potential to improve the quality of health care." International Journal for Quality in Health Care **16**(1): i37-i43.
- Goddard, M., R. Mannion, et al. (2000). "Enhancing performance in health care: a theoretical perspective on agency and the role of information." Health Economics **9**: 95-107.
- Goldacre, M. J., M. Griffith, et al. (2002). "In-hospital deaths as fraction of all deaths within 30 days of hospital admission for surgery: analysis of routine statistics." BMJ **324**: 1069-1070.
- Goldstein, H. and D. J. Spiegelhalter (1996). "League Tables and their Limitations: Statistical Issues in Comparisons of Institutional Performance." Journal of the Royal Statistical Society, Series A **159**: 385-443.
- Gray, A. and S. Harrison, Eds. (2004). Governing medicine: theory and practice. Buckingham, Open University Press.
- Green, J. and N. Wintfeld (1995). "Report Cards on Cardiac Surgeons: Assessing New York State's Approach." New England Journal of Medicine **332**(18): 1229-1232.
- Greenhalgh, T. and et al. (2004). How to spread good ideas: a systematic review. London, National co-ordinating centre for NHS service and delivery and organisation R & D (NCCSDO).
- Gysels, M., R. Hughes, et al. (2004). "What methods do stakeholders prefer for feeding back performance data: a qualitative study in palliative care." International Journal for Quality in Health Care **16**(5): 375-381.
- Harley, M., M. A. Mohammed, et al. (2005). Was Rodney Ledward a statistical outlier? Retrospective analysis using routine hospital data to identify gynaecologists' performance, BMJ, 330:929.
- Harvey, R. (2004). "Using clinical performance information to improve the quality of care in a specialist NHS Trust." Journal of Nursing Management **12**: 427-35.
- Hayes, R. J. (1988). "Methods for assessing whether change depends on initial value." Statistics in Medicine **7**: 915-27.
- Hofer, T. P. and R. A. Hayward (1995). "Can Early Re-Admission Rates Accurately Detect Poor-Quality Hospitals?" Medical Care **33**(3): 234-245.
- Hofer, T. P. and R. A. Hayward (1996). "Identifying Poor-Quality Hospitals: Can Hospital Mortality Rates Detect Quality Problems for Medical Diagnoses?" Medical Care **34**(8): 737-753.
- Ibrahim, J. E. (2001). "Performance indicators from all perspectives." International Journal for Quality in Health Care **13**(6): 431-432.
- Iezzoni, L. I. (1994). "Using Risk-Adjusted Outcomes to Assess Clinical Practice: An Overview of Issues Pertaining to Risk Adjustment." Annals of Thoracic Surgery **58**(6): 1822-1826.
- Iezzoni, L. I. (1997). "The Risks of Risk Adjustment." JAMA **278**(19): 1600-1607.
- Iles, V. and K. Sutherland (2000). Review of existing research evidence on change management and quality improvement. London, National Coordinating Centre for NHS service and delivery and organisation R & D (NCCSDO).
- Jamtvedt, G., Y. J. M, et al. (2005). Audit and feedback: effects on professional practice and health care outcomes (review), The Cochrane Collaboration.
- Kuhlthau, K., T. G. G. Ferris, et al. (2004). "Risk adjustment for pediatric quality indicators." Pediatrics **113**: 210-6.

- Lanier, D. C., M. Roland, et al. (2003). "Doctor performance and public accountability." The Lancet **362**: 1404-1408.
- Leatherman, S. and K. Sutherland (2003). The quest for quality in the NHS: a mid-term evaluation of the ten-year quality agenda. London, The Nuffield Trust.
- Lilford, R., M. A. Mohammed, et al. (2004). "Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma." The Lancet **363**: 1147-1154.
- Mainz, J. (2003). "Developing evidence-based clinical indicators: a state of the art methods primer." International Journal for Quality in Health Care **15**(1): i5-i11.
- Mainz, J. (2004). "Quality indicators: essential for quality improvement." International Journal for Quality in Health Care **16**(1): i1.
- Mannion, R. and H. T. O. Davies (2002a). "Reporting health care performance: learning from the past, prospects for the future." Journal of Evaluation in Clinical Practice **8**(2): 215-28.
- Mannion, R. and H. T. O. Davies (2002b). A Principal-Agent Perspective on Clinical Governance. Getting Health Economics into Practice. D. Kernick. Oxford, Radcliffe Medical Press.
- Mannion, R., H. T. O. Davies, et al. (2005). "Impact of star performance ratings in English acute hospital trusts." Journal of Health Services Research & Policy **10**(1): 18-24.
- Mannion, R. and M. Goddard (2001). "Impact of published outcomes data: case study in NHS hospital trusts." BMJ **323**: 260-3.
- Mannion, R. and M. Goddard (2003). "Public disclosure of comparative clinical performance data: lessons from the Scottish experience." Journal of Evaluation in Clinical Practice **9**(2): 277-86.
- Mannion, R. and M. Goddard (2004). "General practitioners' assessments of hospital quality and performance." Clinical Governance **9**(1): 42-47.
- Mant, J. and N. Hicks (1995). "Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction." BMJ **311**: 793-796.
- Marshall, E. C. and D. J. Spiegelhalter (1998). "Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates." BMJ **316**: 1701-1705.
- Marshall, M. N. and H. T. O. Davies (2001). "Public release of information on quality of care: how are health services and the public expected to respond?" Journal of Health Services Research & Policy **6**(3): 158-162.
- Marshall, M. N. and S. Harrison (2005). "It's about more than money: financial incentives and internal motivation." Quality & Safety in Health Care **14**: 4-5.
- Marshall, M. N., J. Hiscock, et al. (2002). "Attitudes to the public release of comparative information on the quality of general practice care: qualitative study." BMJ **325**: 1278.
- Marshall, M. N., P. S. Romano, et al. (2004). "How do we maximize the impact of the public reporting of quality of care?" International Journal for Quality in Health Care **15**(1): i57-i63.
- Marshall, M. N., P. Shekelle, et al. (2003). "Public reporting on quality in the United States and the United Kingdom." Health Affairs **22**(3): 134-148.
- Marshall, M. N., P. G. Shekelle, et al. (2000). "What do we expect to gain from the public release of performance data? A review of the evidence." JAMA **283**: 1866-74.
- McCarthy, E. P., L. I. Iezzoni, et al. (2000). "Does Clinical Evidence Support ICD-9-CM Diagnosis Coding of Complications?" Medical Care **38**(8): 868-87.
- McGlynn, E. (1998). "Choosing and evaluating clinical performance measures." The Joint Commission Journal on Quality Improvement **24**(9): 470-9.
- McLoughlin, V., S. Leatherman, et al. (2001). "Improving performance using indicators. Recent experiences in the United States, the United Kingdom and Australia." International Journal for Quality in Health Care **13**(6): 455-462.

- Mohammed, M. A., K. K. Cheng, et al. (2001). "Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons." Lancet **357**: 463.
- Mohammed, M. A., J. Raftery, et al. (2004). "Counting in-hospital deaths in England: a comparison of hospital computer systems and mortuary registers." Journal of Health Services Research & Policy **9**(2): 100-03.
- Morton, A. P. (2003). "The use of statistical process control methods in monitoring clinical practice." International Journal for Quality in Health Care **15**(4): 361-362.
- Mukamal, K., D. R. Peterson, et al. (2004). "Variations in risk-adjusted outcomes in a managed acute/long-term care program for frail elderly individuals." International Journal for Quality in Health Care **16**(4): 293-301.
- Naessens, J. M. and T. R. Huschka (2004). "Distinguishing hospital complications of care from pre-existing conditions." International Journal for Quality in Health Care **16**(1): i27-i35.
- Narayn-Lee, M., V. Aylett, et al. (2004). "National clinical indicators: more questions than answers?" British Journal of Health Care Management **10**(2): 44-47.
- Noseworthy, J., G. Ebers, et al. (1994). "The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial." Neurology **44**: 16-20.
- Parry, G. and J. S. Tucker (2004). The effect of the United Kingdom neonatal staffing study results on the prior views of neonatal doctors: a Bayesian analysis. Health care policy, performance and finance: strategic issues in health care management. H. T. O. Davies and M. Tavakoli. Aldershot, Ashgate.
- Powell, A., H. T. O. Davies, et al. (2003). "Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls." Quality & Safety in Health Care **12**: 122-128.
- Pringle, M., T. Wilson, et al. (2002). "Measuring 'goodness' in individuals and healthcare systems." BMJ **325**: 704-707.
- Renshaw, M. and A. Ireland (2003). "Specialty audit leads - has this concept been effective in implementing clinical audit in an acute hospital?" International Journal of Health Care Quality Assurance **16**(3): 136-42.
- Rhydderch, M., G. Elwyn, et al. (2004). "Organisational change theory and the use of indicators in general practice." Quality & Safety in Health Care **13**: 213-7.
- Rosenthal, M. B., R. Fernandopulle, et al. (2004). "Paying for quality: providers' incentives for quality improvement." Health Affairs **23**(2): 127-41.
- Schneider, E. C. and A. M. Epstein (1998). "A Survey of Patients Undergoing Cardiac Surgery." JAMA **279**: 1638-1642.
- Schneider, E. C. and T. Lieberman (2001). "Publicly disclosed information about the quality of health care: response of the US public." Quality & Safety in Health Care **10**: 96-103.
- Scott, I., D. Youlden, et al. (2004). "Are diagnosis specific outcome indicators based on administrative data useful in assessing quality of hospital care?" Quality & Safety in Health Care **13**: 32-39.
- Seneta, E. and J. T. Chen (2005). "Simple Stepwise Tests of Hypotheses and Multiple Comparisons." Internat Statist Rev **73**(1): 21-34.
- Sheaff, R., J. Schofield, et al. (2003). Organizational factors and performance: a review of the literature. London, National Coordinating Centre, Service Delivery and Organisation NHS R&D Programme.
- Sheldon, T. A. (2005). "The healthcare quality measurement industry: time to slow the juggernaut?" Quality & Safety in Health Care **14**: 3-4.
- Shojania, K. G. and J. M. Grimshaw (2005). "Evidence-based quality improvement: the state of the science." Health Affairs **24**(1): 138-50.
- Smith, P. (1995). "On the unintended consequences of publishing performance data in the public sector." International Journal of Public Administration **18**: 277-310.
- Smith, P. (2002). "Performance management in British health care: will it deliver?" Health Affairs **21**(3): 103-115.

- Smith, P. (2004). "Quality incentives: the case of UK general practitioners." Health Affairs **23**(3): 112-118.
- Spiegelhalter, D. J. and A. H. Leyland, Eds. (2001). Multilevel modeling of health statistics. Chichester, John Wiley & Sons.
- Spies, T. H., H. G. Mokkink, et al. (2004). "Which data source in clinical performance assessment? A pilot study comparing self-recording with patient records and observation." International Journal for Quality in Health Care **16**(1): 65-72.
- Steel, N., D. Melzer, et al. (2004). "Developing quality indicators for older adults: transfer from the USA to the UK is feasible." Quality & Safety in Health Care **13**: 260-264.
- Thomas, H. (2003). "Clinical networks for doctors and managers." BMJ **326**: 655.
- Thomas, J. W. and T. P. Hofer (1999). "Accuracy of Risk-Adjusted Mortality Rate As a Measure of Hospital Quality of Care." Medical Care **37**(1): 83-92.
- Thompson, A. and P. Stonebridge (2005). "Building a framework for trust: critical event analysis of deaths in surgical care." BMJ **330**: 1139-1143.
- Thomson, R. G. and J. Lally (2000). "Performance management at the crossroads in the NHS: don't go into the red." Quality in Health Care **9**: 201-202.
- Wensing, M., M. Bosch, et al. (2005). Factors in theories on behaviour change to guide implementation and quality improvement in healthcare. Nijmegen, Centre for Quality of Care Research (WOK).
- Wilcock, P. and R. Thomson (2000). "Modern measurement for a modern health service." Quality in Health Care **9**: 199-200.
- Wilson, A., C. Gibbons, et al. (2004). Surgical wound infection as a performance indicator: agreement of common definitions of wound infection in 4773 patients, BMJ, 329;720-.

You can read and download this document from our website.  
We can also provide this information:

- by email
- in large print
- on audio tape or CD
- in Braille, and
- in community languages.

## **NHS Quality Improvement Scotland**

Edinburgh Office  
Elliott House  
8-10 Hillside Crescent  
Edinburgh EH7 5EA

Phone: 0131 623 4300  
Textphone: 0131 623 4383

Email: [comments@nhshealthquality.org](mailto:comments@nhshealthquality.org)  
Website: [www.nhshealthquality.org](http://www.nhshealthquality.org)

Glasgow Office  
Delta House  
50 West Nile Street  
Glasgow G1 2NP

Phone: 0141 225 6999  
Textphone: 0141 241 6316