

Using performance indicators to improve health care quality in the public sector: a review of the literature

Tim Freeman

Health Serv Manage Res 2002 15: 126

DOI: 10.1258/0951484021912897

The online version of this article can be found at:

<http://hsm.sagepub.com/content/15/2/126>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Health Services Management Research* can be found at:

Email Alerts: <http://hsm.sagepub.com/cgi/alerts>

Subscriptions: <http://hsm.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - May 1, 2002

[What is This?](#)

Using performance indicators to improve health care quality in the public sector: a review of the literature

Tim Freeman

Health Services Management Centre, University of Birmingham, UK

Given the increasing importance of performance indicators in current UK health policy, this paper provides a systematic review of empirical and theoretical writings concerning their use to improve health care quality. The paper outlines potential problems and explores how best to derive, implement and use performance indicator data, presenting results thematically. The two principal uses of indicator systems are as summative mechanisms for external accountability and verification, and as formative mechanisms for internal quality improvement. In the UK, the use of performance indicators in assurance and performance management systems has heavily influenced debate over their value. Major problems reported include the potential to undermine the conditions required for quality improvement, perverse incentives and the difficulty of using data to promote change. Technical problems include indicator selection; the availability, validity and reliability of data; confounding; and problems with robustness, sensitivity and specificity. Factors that help in the derivation, implementation and use of indicator systems include clear objectives, involvement of stakeholders in development, and use of 'soft' data to aid interpretation.

Introduction

Since the early 1980s, there has been an unprecedented increase in the number of health care performance indicators available in the UK National Health Service (NHS). Definitions of performance during this period have varied greatly, and at different points the emphasis has been on economy, efficiency, effectiveness, outputs, 'quality', outcomes, and social results (Talbot, 2000). Consequently, the indicators themselves have also varied widely, in terms of what they purport to measure, their

presentation and their intended audiences. Earlier initiatives, such as the Performance Indicators initiative of 1983, drew on activity and cost data for purposes of internal control by local managers. This emphasis shifted following the publication of *Working for Patients* in 1989, which gave new powers of regulation and performance management to the NHS Management Executive, and required indicators for the purpose of external accountability (Smee, 2000). Thus, the increasing number of indicators at this point partly reflected changing trends in the governance of public organizations away from direct hierarchical control and towards indirect regulatory systems (Jacobs and Manzi, 2000). Such 'hands-off' systems marry central control with local responsibility and seek to verify

Tim Freeman, Health Services Management Centre, Park House, 40 Edgbaston Park Road, Edgbaston, Birmingham B15 2RT, UK.

compliance by auditing performance against targets (Power, 1997; Carter *et al.*, 1992).

New Labour's modernization agenda has again shifted the balance of public sector governance, further extending the role of indirect regulatory systems (Newman, 2000; Goddard *et al.*, 2000a). The White Paper *The new NHS: modern, dependable* (DoH, 1997) outlined a new national Performance Assessment Framework (PAF) as the main instrument for securing management control in the NHS. The PAF marks an attempt to go beyond processes and outputs (Deeming, 2001) and identifies six performance areas: health improvement, fair access, effective delivery, efficiency; patient/carer experience, and health outcomes. Further indicators were announced with the publication of NHS Performance Ratings from September 2001, intended to provide a high level summary of overall performance of all NHS general hospitals, relating principally to patient experience. The trend to increased numbers of indicators shows no signs of slowing, as the number of organizations covered by, and range of indicators included in, the Performance Ratings is set to increase over time (DoH, 2001).

As the scope of performance indicators in the UK NHS continues to widen, it is important that evidence concerning potential difficulties informs their future use. This paper is the first in a series of two, and seeks to summarize the general empirical and theoretical studies of the use of performance indicators to improve health care quality. Having outlined the aims and limitations of such systems, the paper then identifies important lessons for their derivation, implementation and use. The second paper in the series details the development of a measure of cultural climate relating specifically to clinical governance. The intention is to provide information for health-care organizations to use in planning their organizational development agenda.

Methods

Data sources, selection and data extraction

The study comprised an extensive review of the electronic databases Medline, HMIC, ASSIA and BIDS, which were selected for their depth of coverage of health-care management topics. Searches were limited to articles published between January 1985 and July 2001 in

peer-reviewed journals, using a simple truncated search for 'performance indica*'. All identified abstracts were reviewed and full articles retrieved where it was clear that they contained empirical evidence or theoretical discussion on generic issues in the use of performance indicators in assessing health care quality. Where this was not clear from the abstract, full articles were retrieved before making a decision on inclusion or exclusion. Given the focus of the study, disease- and condition-specific papers without general import were excluded, as were papers detailing the reliability and validity testing of specific outcome indicators. Individual study methodologies were not evaluated given their largely discursive nature.

Data synthesis

In total, 125 articles were selected, including reviews, original empirical studies and theoretical discussions. Each article was carefully read and its key findings identified. These findings were classified into three broad categories forming the focus of the review: aims of performance indicators systems; perceived or reported problems; and perceived or reported factors facilitating their derivation, implementation or use. These categories were further refined and organized into the major themes reported in the results section. Although this is a comprehensive review, findings are presented in summary form, illustrating important points. Example references are presented for each theme and a full bibliography is available separately from the author.

Results

A careful reading of the literature reveals two broad traditions: the first is practice-oriented, prescriptive and optimistic of the value and use of performance indicators; the second is more critically engaged and sceptical. These points of approach intersect at important junctures and there is some agreement over what makes indicators desirable and productive and the difficulties involved in their production and use. The central ideal includes selection of appropriate indicators, notions of valid and reliable data collection and sensitivity in the analysis and use of data in order to produce improvements in care quality. Both approaches

are largely discursive, with an underdeveloped empirical base. Empirical contributions mainly comprise single- and multiple-site case studies of the introduction of performance indicator systems and retrospective analyses of routine clinical data. The latter includes the adjustment and modelling of clinical outcome data as performance indicators.

Much of the practice-oriented literature focuses on matters internal to health-care organizations and links conceptually to Total Quality Management (TQM) and Continuous Quality Improvement (CQI) approaches. Some writers make a significant contribution to the type of structures, processes and conditions required to develop performance indicator systems, yet this tradition tends to ignore analytic difficulties and treats the beneficence of performance indicators as axiomatic. It does not consider conceptual problems, nor deal adequately with implementation difficulties. The academic literature focuses on the questions that the practical literature ignores, such as the unintended consequences of performance indicators, and the validity of inferences drawn from data. This literature is more consistently negative, and forms a discourse defined in terms of conceptual difficulties and statistical proof. These concerns are largely the result of the political context of the introduction of performance indicators in the UK public sector, specifically the use of performance league tables.

Aims of performance indicator systems

The two principal uses of indicator systems described in the literature relate to internal and external control and accountability (Davies and Lampel, 1998; Rissel *et al.*, 1998; Smith, 1995b) and formative quality improvement (McGlynn, 1998; Goddard *et al.*, 1999). Boland and Fowler (2000) present a useful matrix structure of performance indicator systems based on two axes: the source of control (internal or external) and the nature of resultant action (positive [supportive and formative] or negative [punitive or summative]). Of the four possible ideal types, the authors contend that the practice of performance management in the UK public sector consists mainly of external, negative approaches. This contrasts with the use of performance indicators espoused by Continuous Quality Improvement (CQI) philosophies,

which may be characterized as internal, positive approaches.

External, summative indicator systems act as mechanisms for external verification of quality improvements and central control (Martin and Kettner, 1997; Pollitt, 1985; Roberts, 1994; Smith, 1990; Smith, 1995a). Such external accountability may be political, commercial or to the community (Leggat *et al.*, 1998) and is ultimately concerned with the renewal of legitimacy (McKevitt, 1993). Information may be useful for purchasers and consumers (McKee and Hunter, 1995; Popovich, 1998) or help to inform political debate (Smith, 1995a). Such systems are consistent with a mode of management concerned with verification of activity, which pushes control into organizational structures, inscribing them within auditable systems. By focusing institutional attention on their performance indicator performance, governments are able to impose a policy agenda on organizations by embedding assumptions related to goals, values and purposes into the selection and structure of indicators (Jacobs and Manzi, 2000).

Those championing performance indicators as a formative mechanism for quality improvement tend to do so from within a CQI paradigm, emphasizing their potential to foster insight into practice (Portelli *et al.*, 1997). Here indicators are used as a focus for feedback and learning, leading to improvement (Schyve, 1995; Tarr, 1995). They act as a vehicle to align the objectives of staff and the organization (Wyszewianski, 1988) as stakeholders discuss and agree which indicators to include.

Carter *et al.* (1992) classify indicators into three types: prescriptive, descriptive and proscriptive. Prescriptive indicators are 'dials' that show achievement against targets, descriptive are 'tin openers' that record changes and proscriptive indicators are 'alarm bells' that specify what should not happen. The two traditions each use indicators in exploratory and confirmatory fashion: acting as 'warning bells' for further investigation (Donabedian, 1992; Holland *et al.*, 1994) or supporting existing suspicion and providing a trigger for action (Birch and Maynard, 1986). Both traditions use benchmarking to identify poor performers and centres of excellence (Nutley and Smith, 1998) making comparisons at the level of clinicians or units (Forster *et al.*, 1990). However, their information requirements and methods differ markedly (Table 1).

Table 1 Differences between accountability and improvement approaches to indicator systems

	Assurance/Accountability	Improvement
Emphasis	Verification and assurance. Measurement oriented	Learning to promote continual improvement. Change oriented
Rationale	Provide external accountability and renew legitimacy	Promote change and improvement in care quality
Culture	Comparisons in order to make summative judgements on care quality. League tables. Blame and shame	Comparisons have a formative emphasis to learn from differences between providers and encourage improvement. Informal benchmarking to promote discussion and change
Precision required	High precision. Use of statistics to identify 'real' differences	Lower precision.
Epistemology	Empirical. Statistical validity and reliability important	Interpretative. Use of other data sources and local information to provide context

Assurance systems tend to use indicators prescriptively, to make comparisons between providers leading to summative judgements on care quality; they seek to prove that providers with poor scores on indicators really are performing badly. This requires very high levels of data precision. The accuracy of information provided by performance indicators is assessed empirically, their claims to objective and material 'truth' settled with recourse to statistical theory. The need for precision generates many problems, specified below. In contrast, indicators intended to provide information for quality improvement are used descriptively. They provide a provisional and partial truth interpreted in the light of local circumstance, and act as a starting point for discussion between stakeholders. While quality improvement models use indicators to develop discussion further, assurance models use them to foreclose it.

Perceived or reported problems with performance indicator systems

In the UK, the use of performance indicators in assurance and performance management systems has heavily influenced debate over their value. Many of the conceptual and technical problems below arise from problems over validity, reliability and perverse incentives that tend to characterize such systems.

Conceptual

In common with all assurance approaches, the primary goal of assurance-focused performance

indicator systems is the verification of improvements; quality must not only improve but be seen to do so. However, faith in measurement may be misplaced. The existence of indicators does not remove the need for trust but relocates it. Instead of relying on the internal control systems of professionals, auditors invest trust in data systems to provide information on practice (Power, 1997). Under such circumstances performance indicator frameworks may simply displace existing informal modes of quality assurance. This is crucial as, because these strategies tend to be informal and unrecorded, there is a risk of assuming that they do not take place. The irony is that new structures may displace these informal strategies and, in verifying the accountability of agents, may generate suspicion and fear, undermining the conditions of trust required for quality improvement (Sitkin and Roth, 1993). The indicator framework may become a ritualistic verification system that does not provide reliable information on activity (Davies, 1998). At its most ineffectual, a performance indicator system may simply provide information to fuel ritualized exchanges between agents and their principals. Deconstructionists characterize such exchanges as cultural performances and describe their publication – as league tables – as a form of contemporary spectacle or theatre (Stronach, 1999).

The use of performance measures is clearly appropriate for monitoring compliance with regulations or comparing realities against a formal plan. However, indicators are of much less use in understanding how interventions or

implementation processes have influenced results. They are not capable of showing why particular results are obtained, which is required to inform policy and programme modifications (Blalock, 1999). While quality-improvement-oriented performance indicator systems may be able to use dialogue between stakeholders to generate insights, this avenue is difficult to pursue in assurance-oriented systems.

Technical

One of the main attractions of indicators is that they promise visible and concrete proof of performance. The argument is that indicators increase objectivity by applying commonly agreed rules of assessment across all organizations, thereby deriving knowledge that is independent of its creators. Yet critics question the very attribute of performance indicators that advocates of assurance-based systems find so appealing: their objectivity. The claim to objectivity is essential to the use of indicators in performance league tables, yet poses many difficulties. Much of the debate is conducted in terms of statistical proof but includes indicator selection, meaningfulness, robustness in the light of adjustments for confounding factors as well as difficulties that arise when using them to inform service change. Indicators based on a limited range of items available in pre-existing information systems may have significant problems with their validity, reliability and comparability.

Indicator selection: Indicators provide information on a potentially limitless number of dimensions and, as indicator systems are unable to capture more than a fragment of what is important about the human experience of health care, some selection is required. A delicate balance needs to be struck between coverage and practicality: too few indicators and important aspects will be missed; too many and the instrument will be impractical to use and costly to maintain. By making some aspects visible, indicators marginalize other aspects and perspectives (Van Peursem *et al.*, 1995). They are thus conceptual technologies (Barnetson and Cutright, 2000), shaping *which* issues are thought about and *how* people think about them.

The structured reporting of indicators further obscures the complexity, contextuality and layered meanings involved in interpretation.

Variation in definitions of what appear to be 'simple' indicators at different centres may lead to the failure to compare like with like (Gross *et al.*, 2000; Jackson, 2001; McColl *et al.*, 2000). Further difficulties in operationalizing indicators arise specifically in the public sector. This is partly due to the lack of a financial 'bottom line', but also because of the existence of multiple conflicting objectives, the over-riding importance of political objectives and short political time-horizons (Hepworth, 1988). There are no technical solutions to these problems and value judgements are required, given the existence of legitimate political debates surrounding the definition of 'appropriate' measures, such as whether to take social factors into account (Stewart and Walsh, 1994).

Data: availability and reliability: There are often problems with the availability of data, resulting in a tendency to focus on measuring what there is data for, rather than items that correlate with the system's goals and objectives (Lowry, 1988; McKee and James, 1997). Data collected by clinicians have the benefit of encouraging discussion and debate for health improvement (Rowan and Black, 2001) but at the potential cost of making external benchmarking difficult (Nyhan and Marlowe, 1995). Data accuracy is also important, the issue being whether between – group differences reflect quality of data rather than quality of care (Kazandjian *et al.*, 1996). Flawed data are likely to increase confrontation and reduce co-operation (McKee and Hunter, 1995), especially when used for summative assurance purposes. All data collection relies on the goodwill of clinicians and is thus susceptible to manipulation, particularly when reward and censure depend on results (Audit Commission, 2000)

Data: validity and confounding: Even where data are available and reliable, they may be potentially misleading and easily misinterpreted (McColl *et al.*, 1998a; Smith, 1995b). Measurement validity reflects the extent to which indicators truly represent a more abstract variable. To be valid measures of health-care quality, indicators need to reflect attributes of the health-care system, rather than attributes of the patient or of other non-health care characteristics. For example, readmission rates are a valid indicator of care quality to the extent that readmission is due

solely to deficiencies in the quality of the previous care. Because measures, and especially outcomes, are affected by other variables, it is difficult to attribute variation to performance (Bull *et al.*, 1994; Leng *et al.*, 1999; Leyland and Boddy, 1997; Milne and Clarke, 1990; Orchard, 1994). The meaning of indicators thus becomes debatable. The central point is that indicators should only relate to factors that are under the control of those under scrutiny (Parry *et al.*, 1998).

To avoid confounding, all other exogenous and endogenous factors affecting the indicators must be controlled (Barnsley *et al.*, 1996; Birch and Maynard, 1996; Davenport *et al.*, 1996). Potential confounding factors include configuration of the local health economy (Brown *et al.*, 1995; Carter, 1989), socio-economic variations (Giuffrida *et al.*, 1999), case mix, co-morbidity and severity (Mant and Hicks, 1996; Rigby *et al.*, 2001). Without adjustment, it is not clear to what extent indicators identify the contribution of health services to health care (Mulligan *et al.*, 2000). There are particular problems associated with outcome measures, as they may occur over long time scales and suffer from problems of attribution, especially in measures of chronic illness (Smith, 1995c). Theoretical models incorporating antecedent factors have been identified as a solution (Ashton and Wray, 1996; Coyle and Battles, 1999) but have data requirements that are unlikely to be provided by routine systems. Data collection costs also need to be considered, as organizations may face significant financial and time costs in producing and reporting the data, potentially incurring significant opportunity costs (Boyne, 2000; Heasel, 2000).

Dealing with confounding: Risk adjustment models may control the effects of confounding. Four principal methods are available, each posing its own difficulties (Blumberg, 1986; Giuffrida *et al.*, 2000). The first is standardization, in which differences in variables such as age and sex mix in a population are standardized. The problem is that there may be other population characteristics that may confound the indicators and we cannot standardize for them unless we know what they are and hold data on them. The second is cluster analysis, which groups units for comparison into clusters sharing similar socio-economic profiles. The problem is

that clustering criteria are arbitrary. The same clusters are often used for all measures, which is plausible only if the variables used in the clustering procedure affect all indicators in the same way. Options three and four are both ways of modelling data. Data Envelope Analysis (DEA) calculates lowest possible indicator rate with given levels of confounding variables and compares an area's actual performance against the best possible rate given its confounding profile. The final option, multiple regression analysis, uses information from all units to predict the indicator rate that a unit should have given the values of its confounding variables. The problem with modelling the data is that residuals that cannot be explained by the model are due to other factors. These may include care quality but could also be due to any other potential variable not in the model. Further, poor practice is included in the norming procedure (Leyland *et al.*, 1991). Perhaps the greatest difficulty is that the two final options also suffer from not being transparent to the end user. The question becomes 'do you trust the statistician', as attempts to increase the validity of the data may simply undermine their credibility. The tendency of different adjustment methods to yield different results brings the robustness of the process further into question (Brown *et al.*, 1995; Nutley and Smith, 1998).

Indicators: robustness, sensitivity and specificity: Indicators may falsely convey an impression of objectivity to what is often weak and ambiguous evidence (Davies and Lampel, 1998; Van Peursem *et al.*, 1995). Small numbers of cases mean low significance (McGlynn, 1998; Sheldon, 1998), the data requirements for precision are excessive (Mant, 1995; Logan, 1991), and random variation in measures may be misinterpreted (Smith, 1995c). The inability to measure social outcomes with precision creates problems for accountability, as comparisons become difficult and potentially deceptive. Year-to-year variation within league table rankings has been high and unreliable and consequently league tables may lead to unnecessary praise or sanction (Wellard, 2000). It is important that indicators are able to identify all poorly performing units (sensitivity) and that all units identified by indicators as performing poorly really are performing poorly (specificity). Poor sensitivity and specificity results in false assurance

or denigration, where indicators incorrectly identify individuals or organizations as poor, or fail to identify the poorly performing (Davies, 1998; Forster *et al.*, 1990; Goldstein and Spiegelhalter, 1996). It should be noted that different levels of sensitivity and specificity are needed depending upon purpose, with summative accountability decisions requiring much higher accuracy than formative developmental ones. Given the disputed nature of their meaning and precision, prescriptive indicators are often used descriptively in practice, to provide a starting point for discussion (Carter, *et al.*, 1992; Linder, 1991). Indeed, the need to take context and other informal 'soft' data into account is widely recognized (Goddard *et al.*, 1999; DoH, 2001).

Indicators: promoting action and change: Indicators have an end-of-process focus, which takes time to collect data and act (Davies and Lampel, 1998; McColl *et al.*, 1998b). They are also a poor motivational device for action and change (Mannion and Goddard, 2001). All indicator systems give rise to perverse incentives and unintended consequences (Goddard *et al.*, 2000b; Roberts, 1994; Smith, 1990; Smith, 1995a). This problem is due to the assumption of a mechanistic response to performance indicator feedback, in which provision of information leads directly to changes in practice. Health care may be better characterized as a complex adaptive system (Plsek *et al.*, 2001) in which people anticipate the reactions of those charged with controlling them. Accordingly, the behaviour of those within a system may give rise to negative unintended consequences (Thompson and Lally, 2000). Smith (1995a) outlines the potential distortions induced by over reliance

on performance indicators by drawing attention to the implicit management incentives of such schemes. Drawing on a wealth of literature, he concludes that the almost universal finding is that performance indicators distort behaviour in unintended ways (Table 2).

When rewards are dependent on data held by professionals, dysfunctional behaviour may result. Potential effects include manipulation of records (McKee and Hunter, 1995) and even changes to clinical practice itself (Smith, 1995b), potentially reducing the number of surgical interventions used on high-risk cases (McKee, 1997). Hannan *et al.* (1994) dispute this, in a secondary analysis of cardiac surgery outcomes showing no such reduction in surgery in high-risk cases.

A further problem with the use of outcome 'rates' as indicators is that they provide no indication of what, if any, action is appropriate to improve health care. They conceal the detail required to show what went wrong (McColl *et al.*, 1998b). The problem is that even if we can make meaningful comparisons and identify deficiencies, it may still be unclear what action to take. This requires CQI-based interventions with clinical teams to improve clinical quality.

Perceived or reported factors facilitating the derivation, implementation and use of performance indicators

Much of the literature on health care performance indicators is concerned with the technical and conceptual difficulties involved. This is unsurprising, given that those developing indicators, rather than those charged with using them, most often write it. However, the big question for practitioners is how to use indicator data most effectively to encourage and develop quality improvement. Those with

Table 2 *Unintended consequences of public sector performance indicator systems (after Smith, 1995a)*

Tunnel vision	Emphasis on phenomena quantified in the measurement scheme
Sub-optimization	Pursuit of narrow local objectives, rather than those of the organization
Myopia	Pursuit of short term targets
Measure-fixation	Pursuit of strategies enhancing the measure rather than the associated objective
Misrepresentation	Deliberate manipulation of data
Misinterpretation	Drawing misleading inferences from raw performance data
Gaming	Deliberate manipulation of behaviour to secure strategic advantage
Ossification	Organizational paralysis due to rigid performance evaluation

responsibility for implementing indicator systems need to bear in mind the need for an analysis of the work environment, design of indicators, integration into the workplace and evaluation and further development of indicators (Cave *et al.*, 1990). While it is rare to conduct an initial analysis of the work environment, it is also critical as it affects the ability of the various interest groups to engage fully with the process. The existence of staff hierarchies or a history of poor experience with quality improvement initiatives will affect the level of staff involvement and support for any initiative. Those charged with implementing work on performance indicators would be wise to identify and address such barriers before developing indicators.

Derivation

It is vitally important that stakeholders share a common understanding of the intended use of proposed indicators. Internal quality improvement and external accountability are distinct requirements and the information and processes required for each differ markedly (Solberg *et al.*, 1997). External assurance purposes require investment in reliable and valid data collection, and modelling to adjust for confounding. Quality improvement approaches require less robust data but require processes that encourage discussion of results and lead to service improvement. A lack of clarity over the aims of an indicator system will inevitably lead to problems over ownership of the data and disputes over their meaning and proper use. It is therefore important that all participants agree on the use of indicator data, and design systems appropriate for the specified task. Before indicators are developed, stakeholders should clearly define their goals (Smith and Frowen, 1997). All principal stakeholders need to be involved in the development of indicators (Goddard *et al.*, 2000a; Nyhan and Marlowe, 1995; Smith, 1995a). This includes clinical, managerial, purchaser and patient perspectives (Popovich, 1998). A number of indicators are required to ensure all dimensions of the area of interest are covered. The need for multiple measures should be balanced against the selection of only those indicators critical to the agreed goals (Leggat *et al.*, 1998).

Finally, indicators are only meaningful if they are markers of outcomes or processes that are

under the influence of clinicians (McKenzie and Shilling, 1998). If clinicians are unable to affect the result, the indicator is redundant.

Implementation

Well-derived indicators can act as a catalyst for change within the organization (Tarr, 1995). While a sensitive development will minimize resistance, managers do need to anticipate resistance during the implementation process (Meekings, 1995). Executives may signal the importance of the initiative by active 'championing' and releasing adequate resources for its successful implementation and delivery (Holloway *et al.*, 1999; Tan, 1999). While executives can point the way, line managers have an important role to play in ensuring top-to-bottom links in the organization (Nyhan and Marlowe, 1995). By using indicators in a transformational manner, managers and clinicians may integrate quality initiatives around an agreed agenda. Indicators should not be used simply as an additional extra but integrated into everyday working practices, as part of both clinical and management processes and the wider organizational culture (Flynn, 1986).

Implementation requires the development of IT systems required for information capture. Wherever possible, automated collection, input, analysis, retrieval and dissemination of data are helpful (Al-Assaf, 1996). The physical provision of IT systems is, of course, only the first step, as data quality (Leggat *et al.*, 1998) and reliability (Schyyve, 1995) require assurance processes.

Use

Given the difficulties involved in the validity of indicators, there is a fine line to be drawn between the use of unadjusted data and attempting to control for confounding factors. Too much adjustment and the immediacy and credibility of indicators is lost; too little adjustment and indicators may be very misleading. Where possible, it is helpful to adjust raw figures for social and environmental factors, plus severity and case mix (Holland *et al.*, 1994; Leyland *et al.*, 1991). However, users should remember that there are no technical solutions to problems of interpretation. Reported indicators are exactly that; indicators to focus attention on issues of interest. They are

neither proof of a problem or its solution. Users would be well advised to offset precision and reliability issues with a holistic impression, derived from a wide range of indicators, and link interpretations to other local sources of knowledge (Goddard *et al.*, 1999; Smith, 1995a). From this perspective, data are sensitizing material only, providing hints to outcomes emerging from activity and subject to debate and discussion (Krivenko and Chodroff, 1994; Smith, 1995c). Much of the debate around statistical significance of results can be avoided, as adverse occurrences in small numbers of cases require individual case analysis to explore circumstances, not aggregation into population rates (Logan, 1991).

To be consistent with the above, participants require learning-focused, non-judgmental feedback (Greene, 1999; Nyhan and Marlowe, 1995; Tarr, 1995). Indicators should be used to learn and correct rather than blame, with room for discussion and management judgement. Ideally, multidisciplinary 'performance improvement teams' should discuss identified problems (Kazandjian *et al.*, 1996) and agree planned changes in the light of their deliberations. To develop cooperation, users would be wise to avoid external release of indicators designed for internal quality improvement purposes. The control of data should rest with those responsible for service delivery and improvement, to avoid undermining the trust necessary for quality improvement. Finally, goals, indicators and their underlying values require constant re-evaluation to ensure continued relevance to the organization (Portelli *et al.*, 1997).

Conclusions

The literature describes two principal uses of indicator systems: as a summative mechanism for external accountability and verification in assurance systems, and as a formative mechanism for internal quality improvement. The weight of evidence considered in this review suggests that the use of performance indicators in a summative way as a basis for praise or sanction is almost inevitably corrosive and corrupting of the indicators themselves. Such accounting systems place trust in systems rather than individuals, further undermining the conditions of trust required for quality improvement. A range of technical problems

arises due to the precision of data required to make summative comparisons and further negative unintended consequences follow from the pressures on clinicians and managers to 'get good results'. Of far greater potential benefit is the formative use of indicators as clues to performance, discussed and interpreted by clinicians and managers in the light of local contexts and with the aim of continuously improving the quality of clinical care. Such approaches foster trust and communication between clinicians and managers, with the result that they are better able to work through problems with care delivery and improve quality.

It is clear that indicators of health care quality are not axiomatically good. They may be inaccurate, misleading and even dangerous in their negative unintended consequences. They are very seductive as they promise an objective view of health care quality, yet it is a promise that they are ultimately unable to keep. Those using indicators of health care quality should do so with caution and remember that they are precisely that – indicators, for further investigation and discussion, requiring cautious interpretation in the light of local circumstances. It is encouraging that documentation outlining the PAF explicitly accepts the need for such cautious interpretation. Failure to do so in practice will generate strong perverse incentives and the pressure on clinicians and managers to 'game' the system will be very difficult to resist.

It is also clear that the NHS can expect increased national comparisons based on performance indicators in the future. There is a danger of overplaying both the objective nature of summary measures of health care quality and the usefulness of centrally collated data in encouraging improvement. The intention behind the NHS Performance Ratings is to gain an overall view of the performance of trusts, by aggregating indicators. The system extends the use of performance indicators beyond accountability and into future resource allocation. Classification will have direct effects on organizations; those fortunate enough to be graded as 'three star' will have considerable freedom to develop local services; others will have less freedom and will receive scrutiny and developmental advice from regional offices or the Modernisation Agency.

The danger is that we may be embarking on another round in the 'natural cycle' of assurance-based indicator systems (Cave *et al.*, 1990). The cycle starts when a crisis flags the need for

accountability. This triggers new indicators and a drive for improvement. To maintain improvement, managers have an incentive to manipulate ('game') the indicators. An attempt to correct this via a proliferation of indicators leads to a complex system that is difficult to understand and then falls into misuse as attention shifts within the organization. There is potentially much to be gained by the sensitive and cautious use of performance indicators and much to be lost if they are used without due caution.

References

- Al-Assaf, A. F. Healthcare outcomes management and quality improvement. *Journal of the Royal Society of Health* 1996; **116**(4): 245–51
- Ashton, C. M. and Wray, N. P. A conceptual framework for the study of early admission as an indicator of quality of care. *Social Science and Medicine* 1996; **43**(11): 1533–41
- Audit Commission. *Aiming to Improve the Principles of Performance Measurement*. London: Audit Commission, 2000
- Barnetson, B. and Cutright, M. Performance indicators as conceptual technologies. *Higher Education* 2000; **40**: 277–92
- Barnsley, J., Lemieux-Charles, L. and Baker, G. R. Selecting clinical outcome indicators for monitoring quality of care. *Healthcare Management Forum* 1996; **9**(1): 5–12
- Birch, S. and Maynard, A. Performance indicators and performance assessment in the UK National Health Service: implications for management and planning. *International Journal of Health Planning and Management* 1986; **1**: 143–56
- Blalock, A. B. Evaluation research and the performance management movement: from estrangement to useful integration. *Evaluation* 1999; **5**(2): 117–49
- Blumberg, M. Risk adjusting health care outcomes: A methodologic review. *Medical Care Review* 1986; **43**(2): 351–93
- Boland, T. and Fowler, A. A systems perspective of performance management in public sector organisations. *The International Journal of Public Sector Management* 2000; **13**(5): 417–46
- Boyne, G. External regulation and best value in local government. *Public Money and Management* 2000; **20**: 7–12
- Brown, R. B., McCartney, S. and Bell, L. Why the NHS should abandon the search for the universal outcome measure. *Health Care Analysis* 1995; **3**: 191–5
- Bull, A. R., Hatton, P., Bensley, D. C., Bull, S. J. and Fryers, P. T. Standardised mortality from cervical cancer: a measure of performance? *Journal of Public Health Medicine* 1994; **16**(1): 16–22
- Carter, N. Performance indicators: backseat driving or hands off control? *Policy and Politics* 1989; **17**(2): 131–8
- Carter, N., Klein, R. and Day, P. *How Organizations Measure Success: the Use of Performance Indicators in Government*. London: Routledge, 1992
- Cave, M., Kogan, M. and Smith, R. *Output and Performance Measurement in Government: the State of the Art*. London: Jessica Kingsley Publishers, 1990
- Coyle, Y. M. and Battles, J. B. Using antecedents of medical care to develop valid quality measures. *International Journal for Quality in Health Care* 1999; **11**(1): 5–12
- Davenport, R. I., Dennis, M. S. and Warlow, C. P. Effect of correcting outcome data for case mix: an example from stroke medicine. *BMJ* 1996; **312**: 1503–5
- Davies, H. T. O. Performance management using health outcomes: in search of instrumentality. *Journal of Evaluation in Clinical Practice* 1998; **4**(4): 359–62
- Davies, H. T. O. and Lampel, J. Trust in performance indicators? *Quality in Health Care* 1998; **7**: 159–62
- Deeming, C. Performance indicators: making information out of data. *Health Care UK* 2001; spring: 33–41
- Department of Health. *The New NHS: Modern, Dependable*. London: HMSO, 1997
- Department of Health. *NHS Performance Ratings: Acute Trusts 2000/01*. London: HMSO, 2001
- Donabedian, A. The role of outcomes in quality assessment and assurance. *Quality Review Bulletin* 1992; **18**(11): 56–60
- Flynn, N. Performance measurement in public sector services. *Policy and Politics* 1986; **14**(3): 389–404
- Forster, D. P., Frost, C. E. B. and Morris, D. Modifying practice: Two initiatives in the English National Health Service. *Journal of Public Health Policy* 1990; **11**(1): 81–105
- Giuffrida, A., Gravelle, H. and Roland, M. Measuring quality of care with routine data: avoiding confusion between performance indicators and health outcomes. *BMJ* 1999; **319**: 94–8
- Giuffrida, A., Gravelle, H. and Roland, M. Performance indicators for managing primary care: The confounding problem. In: P. Smith (ed.) *Reforming Markets in Health Care: an Economic Perspective*. Buckingham: Open University Press, 2000: 162–86
- Goddard, M., Mannion, R. and Smith, P. C. Assessing the performance of NHS hospital trusts: the role of 'hard' and 'soft' data. *Health Policy* 1999; **48**(2): 119–34
- Goddard, M., Mannion, R. and Smith, P. Enhancing performance in health care: a theoretical perspective on agency and the role of information. *Health Economics* 2000a; **9**: 95–107
- Goddard, M., Mannion, R. and Smith, P. The performance framework: taking account of economic behaviour. In: P. Smith (ed.) *Reforming Markets in Health Care: an Economic Perspective*. Buckingham: Open University Press, 2000b: 138–61
- Goldstein, H. and Spiegelhalter, D. J. League tables and their limitations: statistical issues in

- comparisons of institutional performance. *Journal of the Royal Statistical Society A* 1996; **159**: 385–443
- Greene, J. C. The inequality of performance measurements. *Evaluation* 1999; **5**(2): 160–72
- Gross, P. A., Braun, B., Kritchevsky, S. B. and Simmons, B. P. Comparison of clinical indicators for performance measurement of health care quality: a cautionary note. *British Journal of Clinical Governance* 2000; **5**(4): 202–11
- Hannan, E. L., Kilburn, H., Racz, M., Shields, E. and Chassin, M. R. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA* 1994; **271**(10): 761–6
- Heasel, S. The cost of performance indicators. *Health Care Risk Report* 2000; **6**(4): 21–2
- Hepworth, N. P. Measuring performance in non-market organizations. *International Journal of Public Sector Management* 1988; **1**(1): 16–26
- Holland, W. W., Fitzgerald, A. B., Hildrey, S. J. and Phillips, S. J. Heaven can wait. *Journal of Public Health Medicine* 1994, **16**(3): 321–30
- Holloway, I., Francis, G. and Hinton, M. A vehicle for change? A case study of performance improvement in the 'new' public sector. *The International Journal of Public Sector Management* 1999; **12**(4): 351–65
- Jackson, A. An evaluation of evaluation: problems with performance measurement in small business loan and grant schemes. *Progress in Planning* 2001; **55**: 1–64
- Jacobs, K. and Manzi, T. Performance indicators and social constructivism: conflict and control in housing management. *Critical Social Policy* 2000; **20**(1): 85–103
- Kazandjian, V. A., Thomson, R. C., Law, W. R. and Waldron, K. Do performance indicators make a difference? *The Joint Commission Journal of Quality Improvement* 1996; **22**(7): 482–91
- Krivenko, C. and Chodroff, C. The analysis of clinical outcomes: getting started in benchmarking. *The Joint Commission Journal on Quality Improvement* 1994; **20**(5): 260–6
- Leggat, S. G., Narine, L., Lemieux-Charles, L., Barnsley, J., Baker, G. R., Scotte, C. *et al.* A review of organizational performance assessment in health care. *Health Services Management Research* 1998; **11**: 3–23
- Leng, G. C., Walsh, D., Fowkes, F. G. R. and Swainson, C. P. Is the emergency readmission rate a valid outcome indicator? *Quality in Health Care* 1999; **8**: 234–8
- Leyland, A. H., Boddy, F. A. Measuring performance in hospital care: Length of stay in gynaecology. *European Journal of Public Health* 1997; **7**: 136–43
- Leyland, A. H., Pritchard, C. W., McLoone, P. and Boddy, F. A. Measures of performance in Scottish maternity hospitals. *BMJ* 1991; **303**: 389–93
- Linder, J. C. Outcomes measurement: compliance tool or strategic initiative? *Health Care Management Review* 1991; **16**(4): 21–33
- Logan, S. Outcome measures in child health. *Archives of Disease in Childhood* 1991; **66**: 745–8
- Lowry, S. Focus on performance indicators. *BMJ* 1998; **296**: 992–4
- McColl, A., Gabbay, I. and Roderick, P. Improving health outcomes—a review of case studies from English health authorities. *Journal of Public Health Medicine* 1998a; **20**(3): 302–11
- McColl, A., Roderick, P., Gabbay, I., Smith, H. and Moore, M. Performance indicators for primary care groups: an evidence based approach. *BMJ* 1998b; **317**: 1354–60
- McColl, A., Roderick, P., Smith, H., Wilkinson, E., Moore, M. Exworthy, M. *et al.* Clinical governance in primary care groups: the feasibility of deriving evidence-based performance indicators. *Quality in Health Care* 2000; **9**: 90–7
- McGlynn, E. A. Choosing and evaluating clinical performance measures. *The Joint Commission Journal on Quality Improvement* 1998; **24**(9): 470–9
- McKee, M. Measuring hospital performance: are we asking the right questions? *Journal of the Royal Society of Medicine* 1997; **90**: 187–91
- McKee, M. and Hunter, D. Mortality league tables: do they inform or mislead? *Quality in Health Care* 1995; **4**: 5–12
- McKee, M. and James, P. Using routine data to evaluate quality of care in British hospitals. *Medical Care* 1997; **35**(10) Suppl.: OS102–OS111
- McKenzie, F. C. and Shilling, M. D. Avoiding performance measurement traps: Ensuring effective incentive design and implementation. *Compensation and Benefits Review* 1998; **30**(4): 57–65
- McKevitt, D. Managing for results: the role of performance measurement in health care delivery. *Administration* 1993; **41**(3): 307–22
- Mannion, R. and Goddard, M. Impact of published clinical outcomes data: case study in NHS hospital trusts. *BMJ* 2001; **323**: 260–3
- Mant, J. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 1995; **311**: 793–6
- Mant, J. and Hicks, N. Health status measurement and the assessment of medical care. *International Journal for Quality in Health Care* 1996; **8**(2): 107–9
- Martin, L. L. and Kettner, P. M. Performance measurement: the new accountability. *Administration in Social Work* 1997; **21**(1): 17–29
- Meekings, A. Unlocking the potential of performance measurement: A practical implementation guide. *Public Money and Management* 1995; **15**(4): 5–12
- Milne, R. and Clarke, A. Can readmission rates be used as an outcome indicator? *BMJ* 1990; **301**: 1139–40
- Mulligan, J., Appleby, J. and Harrison, A. Measuring the performance of health systems: indicators still fail to take socio-economic factors into account. *BMJ* 2000; **321**: 191–2
- Newman, J. Beyond the new public management? Modernising public services. In: J. Clarke, S. Gewirtz and E. McLaughlin (eds) *New Managerialism, New Welfare?* London: Sage, 2000: 45–61

- Nutley, S. and Smith, P. C. League tables for performance improvement in health care. *Journal of Health Services Research and Policy* 1998; 3(1): 50–7
- Nyhan, R. C. and Marlowe, H. A. Performance measurement in the public sector: Challenges and opportunities. *Public Productivity and Management Review* 1995; 18(4): 333–48
- Orchard, C. Comparing health outcomes. *BMJ* 1994; 308: 1493–6
- Parry, G. J., Gould, C. R., McCabe, C. J. and Tarnow-Mordi, W. O. Annual league tables of mortality in neonatal intensive care units: longitudinal study. *BMJ* 1998; 316: 1932–5
- Plsek, R. and Greenhalgh, T. Complexity science: the challenge of complexity in health care. *BMJ* 2001; 323: 625–8
- Pollitt, C. Measuring performance: A new system for the National Health Service. *Policy and Politics* 1985; 13(1): 1–15
- Popovich, J. A. Multidimensional performance measurement. *Journal of Nursing Care Quality* 1998; 12(4): 14–21
- Portelli, B., Williams, J. and Collopy, B. Using clinical indicators to change clinical practice. *Journal of Quality in Clinical Practice*. 1997; 17: 195–202
- Power, M. *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press, 1997
- Rigby, K. A., Palfreyman, S. and Michaels, J. A. Performance indicators from routine hospital data: death following aortic surgery as a potential measure of quality of care. *British Journal of Surgery* 2001; 88: 964–8
- Rissel, C., Holt, P. and Ward, J. Applying a health outcomes approach in a health service unit. *Australian Health Review* 1998; 12(1): 168–81
- Roberts, A. Integrating strategy with performance measures. *Management Development Review* 1994; 7(6): 13–5
- Rowan, K. and Black, N. A bottom-up approach to performance indicators through clinician networks. *Health Care UK* 2001; spring: 42–6
- Schyve, P. Models for relating performance measurement and accreditation. *International Journal of Health Planning and Management* 1995; 10(3): 231–41
- Sheldon, T. Promoting health care quality: what role performance indicators? *Quality in Health Care* 1998; 7(Suppl): s45–s50
- Sitkin, S. B. and Roth, N. L. Explaining the limited effectiveness of legalistic 'remedies' for trust/distrust. *Organizational Science* 1993; 4(3): 367–92
- Smee, C. The performance assessment framework: where did it come from and where is it going? *Health Care UK* 2000; spring: 3–8
- Smith, H. and Frowen, B. Assessing success and performance within the public sector. *The Mental Health Review* 1997; 2(3): 8–15
- Smith, P. The use of performance indicators in the public sector. *Journal of the Royal Statistical Society A* 1990; 153(1): 53–72
- Smith, P. The unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration* 1995a; 18(2): 277–310
- Smith, P. Outcomes related performance indicators and organisational control in the public sector. In: J. Holloway and G. Mallory (eds) *Performance Measurement and Evaluation*. London: Sage, 1995b
- Smith, P. Performance indicators and outcome in the public sector. *Public Money and Management* 1995c; 15(4): 13–6
- Solberg, L. J., Mosser, G. and McDonald, S. The three faces of performance measurement: Improvement, accountability and research. *Journal on Quality Improvement* 1997; 23(3): 136–47
- Stewart, J. and Walsh, K. Performance measurement: when performance can never be finally defined. *Public Money and Management* 1994; 14: 45–9
- Stronach, I. Shouting theatre in a crowded fire: educational effectiveness as cultural performance. *Evaluation* 1999; 5(2): 173–93
- Talbot, C. Performing 'performance'—a comedy in five acts. *Public Money and Management* 2000; 20(4): 63–8
- Tan, J. K. H. The critical success factor approach to strategic alignment: seeking a trail from a health organizations goals to its management information infrastructure. *Health Services Management Research* 1999; 12: 246–57
- Tarr, J. D. Performance measurements for a continuous improvement strategy. *Hospital Material Management and Quality* 1995; 18(12): 77–85
- Thompson, R. G. and Lally, J. Performance management at the crossroads in the NHS: don't go into the red. *Quality in Health Care* 2000; 9(1): 201–2
- Van Peurse, K. A., Pratt, M. J. and Lawrence, S. R. Health management performance: a review of measures and indicators. *Accounting, Auditing & Accountability Journal* 1995; 8(5): 34–70
- Wellard, S. *Community Care* 2000; 26 Oct–21 Nov: 10–11
- Wyszewianski, L. The emphasis on measurement in quality assurance: reasons and implications. *Inquiry* 1988; 25: 424–36