Process versus outcome indicators in the assessment of quality of health care

JONATHAN MANT

Department of Primary Care & General Practice, Medical School, University of Birmingham, Birmingham, UK

Abstract

This paper reviews the relative strengths and weaknesses of outcome and process measures as performance indicators in health care. Differences in outcome may be due to case mix, how the data were collected, chance, or quality of care. Health care is only one determinant of health and other factors have important effects on health outcomes, such as nutrition, environment, lifestyle and poverty. The advantages of process measures are that they are more sensitive to differences in the quality of care and they are direct measures of quality. However, outcome measures are of greater intrinsic interest and can reflect all aspects of care, including those that are otherwise difficult to measure such as technical expertise and operator skill. Outcome indicators can be improved if efforts are made to standardize data collection and case mix adjustment systems are developed and validated. It is argued that this is worth doing only where it is likely that variations in health care might lead to significant variations in health outcome and where the occurrence of the outcome is sufficiently common that the outcome indicator will have the power to detect real differences in quality. If these conditions are not met, then alternative strategies such as process measurement and risk management techniques may be more effective at protecting the public from poor quality care.

Key words: clinical competence, medical audit, outcome and process measurement (health care), quality assurance

High-profile failures of health care in the UK have kept the spotlight on methods to monitor the quality of health services. Hot on the heels of the 'Bristol case', a 'once in a lifetime drama' [1] in which three doctors were found guilty by the General Medical Council of 'serious professional misconduct' in relation to the deaths of 29 babies and young children undergoing paediatric cardiac surgery, came the 'Shipman case' [2]. Shipman was a British General Practitioner who was convicted of murdering 15 of his patients, although he is suspected of murdering many more. A recent analysis suggests that monitoring outcome could have identified significant deviations from expected mortality rates in both these examples [3]. Others have argued that taking such an approach might distract from more mundane methods that nevertheless may be more effective at protecting patients against poor care [4]. The debate over what are the most useful types of performance indicator to monitor the quality of health care is of international concern [5]. It is simplistic to view process and outcome measures as being in competition with each other, but there are circumstances where one type of measure is likely to be more useful than the other. The aim of this paper is to review the relative strengths and weaknesses of outcome measures compared with process measures as performance indicators for health care and so to consider in which circumstances it might be appropriate to use outcome measures to monitor health care. In order to do this, it is important to consider the different contexts in which this debate is taking place.

A question of perspective

An important contextual issue is the purpose for which the performance indicator is to be used and by whom. The aim of using performance indicators may be to:

- inform policy making or strategy at a regional or national level,
- improve the quality of care of a health care facility,
- monitor performance of health care funders such as HMOs,
- identify poor performers to protect public safety,
- provide consumer information to facilitate choice of health care provider.

At the national level, performance indicators can be used in two ways; to compare performance between countries and to inform policy within a country. The summary measure of population health developed by the World Health Organization, disability-adjusted life expectancy (DALE), allows direct comparison of the health of countries, which can raise

Address reprint requests to J. Mant, Department of Primary Care & General Practice, Medical School, University of Birmingham, Birmingham, B15 2TT, UK, E-mail: j.w.mant@bham.ac.uk

© 2001 International Society for Quality in Health Care and Oxford University Press
important health policy questions, though methodological concerns currently limit the usability of the data [6,7]. Within a country, use of population outcome data can help formulate and monitor health policy. Health care is only one determinant of health. Historically, other factors, such as nutrition, environment, lifestyle, poverty and social structure of society have been demonstrated to have powerful effects on health as measured by mortality rates [8,9]. Contemporary data suggest that these factors continue to play an important role. For example, changes in lifestyle, in particular diet and smoking, appear to account for two-thirds of the 31% decline in coronary heart disease that occurred between 1980 and 1992 in US women [10]. Similarly, changes in the prevalence of risk factors appear to have accounted for half of the reduction in coronary heart disease mortality that has been observed in recent years in countries such as Scotland and New Zealand [11,12]. Setting mortality targets can encourage policy makers and government to consider policies and actions outside the direct confines of the health care system that might have an impact on health. For example, the mortality targets set by the UK government for diseases such as heart disease, stroke and cancer were accompanied by an explicit inter-sectoral strategy for health. This emphasized the importance of non-health care factors as diverse as the restriction of sale of cigarettes to underage smokers, the cost of fruit and vegetables and unemployment [13].

A narrower perspective is obtained if one takes as one's population of interest those who are admitted to hospital. For example, 70% of deaths from acute myocardial infarction occur in people who never reach hospital [14]. This moves the agenda from disease prevention (whether by agencies within or outside the health service) to disease treatment. This is relevant to bodies such as Health Maintenance Organizations (HMOs) in the US or primary care groups and trusts in the UK who, as part of their remit, need to monitor the quality of the hospital services that they procure. This is also an appropriate breadth of perspective for health care providers seeking to improve the quality of their own care through such mechanisms as clinical audit. This perspective is also of relevance to the public, whether to promote market competition in privately funded health services or public accountability of state-financed health care [15].

In general, the broader the perspective required, the greater the relevance of outcome measures, since they reflect the inter-play of a wide variety of factors, some directly related to health care, others not. As the perspective narrows, to hospitals, or to departments, or indeed to individual doctors, outcome measures become relatively less and process measures relatively more useful. In order to consider why this is the case, it is necessary to explore the possible causes of variation in outcome between health care providers.

What causes variation in outcome between health care providers?

There are four major categories of explanation that need to be considered, as shown in Table 1. The first of these is whether observed differences might be due to differences in the type of patient that are cared for by the different providers. Factors such as age, gender, co-morbidity, severity of disease and socio-economic status would come under consideration here. The importance of this cause of variation is illustrated by studies where differences in crude outcome rates disappear when the rates are adjusted to take account of these confounding factors. For example, the introduction of a stroke unit in Edinburgh, Scotland, was associated with a 40% reduction in mortality of stroke patients admitted to hospital. However, this difference disappeared once the results were adjusted for case mix [16].

A second cause of variation in outcome is differences in the way data are collected. An outcome indicator used as a performance indicator has different component parts. For example, if the outcome measure is a dichotomous variable such as death, these comprise a numerator, a denominator and case-mix data used to adjust for potential confounding. Differences in the measurement of any of these will lead to apparent differences in outcome. For example, an important factor in explaining an observed decline in case-mix adjusted mortality from cardiac surgery in New York State is an apparent increase in the prevalence of risk factors in the patients upon whom surgery was performed [17]. Between 1989 and 1991, there was an increase in the reported prevalence of renal failure, congestive heart failure, chronic obstructive pulmonary disease and unstable angina in patients undergoing cardiac surgery. These increases are likely to have reflected changes in how these risk factors were ascertained, rather than genuine changes in case mix. If this is the case, then this would account for 40% of the observed reduction in case-mix adjusted mortality.

Thirdly, observed differences may be due to chance. For example, in the European Carotid Surgery Trial, the risk of major stroke or death within 30 days of carotid endarterectomy was 7% (95% confidence interval, 5.8–8.3). One hundred and forty-seven surgeons took part in this trial and surgeon-specific mortality was in the range of 0–50%. However, because of the small number of patients included in the trial by each surgeon, one cannot conclude either that the surgeons with no operative strokes or deaths were better than the rest, or that the surgeon with a 50% complication rate was worse, as these differences might have arisen by chance [18]. Random variation is influenced both by number of cases included and the frequency with which the outcome occurs. For example, to detect a 30% difference in outcome between two units performing carotid endarterectomy with 80% power at a significance level of 5%, with one unit achieving a 7% death and complication rate and another unit a 10% rate, would require the audit of 1422 carotid endarterectomies in each unit. Given that hospitals in New York State each performed on average 50 carotid endarterectomies a year in 1995 [19], such a difference is unlikely to be detected. Conversely, to detect a 30% difference in outcome between units treating patients with myocardial infarction, with one unit experiencing a 30% mortality and the other unit a 21% mortality, would only require the audit of 389 patients in each unit [20]. Because this number of
Table 1 Reasons for variations in outcome

<table>
<thead>
<tr>
<th>Category of explanation</th>
<th>Sources of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differences in type of patient</td>
<td>Confounding by patient characteristics</td>
</tr>
<tr>
<td>Differences in measurement</td>
<td>Ascertainment and definition of cases, outcome and risk factors</td>
</tr>
<tr>
<td>Chance</td>
<td>Random variation, influenced by number of cases and frequency with which outcome occurs</td>
</tr>
<tr>
<td>Differences in quality of care</td>
<td>Use of proven interventions</td>
</tr>
</tbody>
</table>

patients would be admitted in less than a year to the typical hospital, detecting such a difference is feasible.

Finally, differences in outcome may reflect genuine differences in quality of care. This may be due to variations in the use of interventions, such as use of aspirin and beta-blockers in acute myocardial infarction [21], or less measurable aspects such as the skill of a surgeon or of a cardiac arrest team. It is these possible causes of variation that are of relevance if an outcome measure is to be used as a performance indicator. In this circumstance, the conclusion that a variation in outcome is due to a difference in quality of care is essentially a diagnosis of exclusion. If one cannot explain the variation in terms of differences in the type of patient, in how the data were collected, or in terms of chance, then quality of care becomes a possible explanation. However, the conclusion that differences in outcome are due to differences in quality of care will always be tentative and open to the possibility that the apparent association between a given unit and poor outcome is due to confounding by some other factor that was not measured, or measured inadequately.

Advantages of process measures

An important consideration is the extent to which an indicator will detect a genuine difference in quality. An intrinsic advantage of process measures is that they are more sensitive than outcome measures to differences in the quality of care. For example, if real differences in the quality of care arising from differential use of proven interventions between two hospitals treating patients with myocardial infarction resulted in a 10% relative difference in hospital-specific mortality rates (30% versus 27%), it would require following up 3619 patients in each hospital to detect such a mortality difference and be confident that it was not due to chance (with 80% power at a significance level of 5%) [20]. In a typical UK hospital, this would need 8 years of data collection. However, if one measured the uptake of the proven interventions that would lead to such a mortality difference, it would only be necessary to obtain data on the care of 48 patients in each hospital to detect a significant difference in the process of care [20]. Similarly, one cannot conclude that the lack of difference in mortality of patients admitted to hospital with stroke in Edinburgh after the introduction of a stroke unit reflects a failure of the stroke unit [16]. To be confident that the stroke unit had failed in bringing about a reduction in mortality that would be consistent with the trial evidence would have required a 10 year study [22].

A second advantage of process measures is that they are easy to interpret. A process measure such as use of aspirin in acute myocardial infarction is a direct measure of quality, whereas hospital-specific mortality from myocardial infarction is only an indirect measure. As discussed above, if differences in outcome are observed, then alternative explanations need to be considered before one can conclude that the difference reflects true variations in the quality of care. Conversely, a process measure is straightforward to interpret: the more people without contra-indications who receive a proven therapy, the better. A consequence of this is that the necessary remedial action is clearer (use the therapy more often). However, if one does conclude that a higher mortality rate is due to poor quality of care, it is not immediately obvious what action needs to be taken, unless perhaps audits of the process of care have also been undertaken in parallel. Depending upon perspective (see above), this may not be a problem. If the information is to be used by purchasers or consumers of health care to influence choice of provider (and so improve market efficiency), then the underlying cause of the differences in quality of care is arguably of less importance.

Outcome measures do, however, have a role in the monitoring of the quality of health care, in that they do have some intrinsic advantages.

Advantages of outcome measures

One attraction of outcome measurement is that it is a measure of something that is important in its own right. It is interesting to know that the death rate from myocardial infarction varies from hospital to hospital, even if the reasons for the differences have nothing to do with the quality of care. A
process measure is only of value if it is assumed to have a link to outcome. By itself, the process measure is of little intrinsic interest.

Secondly, outcome measurement will reflect all aspects of the processes of care and not simply those that are measurable or measured. Attributes such as technical expertise and operator skill, while likely to be important determinants of outcome in some situations, cannot easily be captured as performance indicators. Where you how do something is as important as what you do, process measures will be unable to capture the distinction [23]. Thus, where technical skill is relatively unimportant (e.g. giving an aspirin tablet), then the process measure is satisfactory. Where the technical skill is important (e.g. perforating coronary angioplasty or carotid endarterectomy), it not only matters that the procedure is performed on the correct patients (a process measure), but also how well the procedure was carried out.

Another possible reason why outcome indicators are often used in some countries is that the data, at least to construct simple rates, are available from routine information systems. Hospitals will routinely record the admission diagnosis and other data that can be used for simple case mix adjustment, such as age, gender and place of residence. In the UK, process data has been less readily available in the past and has required retrospective case note reviews. This is not an issue in countries such as the US where process data are required to support the construction of charges for health care. As the importance of process indicators becomes recognized, prospective process data are more likely to be recorded in other countries. For example, in the UK, the government's National Service Framework for Coronary Heart Disease specifies process measures that health care providers need to record and on which their performance will be monitored [24].

**When to use outcome indicators?**

The major problem with outcome measures as performance indicators are that they are not a direct measure of quality of health care in the same way that process measures are. A patient admitted to hospital with a heart attack may receive atrocious care, yet despite this, is likely to survive [25]. However, steps can be taken to minimize the possibility of a false conclusion being drawn about the quality of care based on outcome measurement. Including sufficient numbers of patients will reduce the possibility of random variation masking real differences or making spurious differences appear. Statistical significance testing can also minimize the possibility of the latter type of error. Standardising how data are collected can reduce the extent to which differences in measurement can be the cause of observed variation. Development of sophisticated case mix adjustment systems can reduce the possibility that observed differences are due to differences in the types of patient that are treated. The question is in what circumstances is it worth going to the effort and expense of setting up an outcome monitoring system covering adequate numbers of patients using consistent methods of case definition and ascertainment and including sufficient case mix data to allow risk-adjusted rates to be derived? (Given that process measures offer an easier to interpret and less costly alternative).

The first issue to consider is the extent to which variations in quality of care might lead to significant variations in outcome. Taking stroke care as an example, a reasonable estimate of the impact that good quality care might have on outcome is provided by the overview of trials looking at the effect of organized inpatient care for stroke [26]. This concluded that stroke unit care was associated with a 17% reduction in the odds of death after 1 year. Observed variations in outcome following hospital admission with stroke are much greater than this [27]. This suggests that the bulk of the observed variation in mortality is likely to be due to differences in case mix than to differences in quality of care. If large differences in mortality persist after case mix adjustment, then the likeliest explanation is that the case mix adjustment was inadequate. In such circumstances, it is probably not worth spending resource on developing case mix adjustment systems (and standardising the collection of all the additional data that this would entail) to refine the outcome indicator, since it is unlikely to be able to detect real differences in quality. However, in circumstances where one might anticipate that quality of care was a substantial determinant of outcome, then outcome measures are likely to be of more value as an indicator of quality. For example, mortality following surgery is likely to depend to a significant extent upon the skill and technique of the surgeon. Indeed, for some operative procedures, the benefits of the surgery are outweighed by the potential harms if the complication rate is high: a review of the evidence on carotid endarterectomy concluded that it is effective only if performed by surgeons who have complication rates of less than 6% [28]. In such instances, it is likely to be of value to use case mix adjustment systems [29], since residual variation is more likely to reflect quality of care.

The second issue to take into account is whether the outcome indicator is likely to have the statistical power to detect differences in quality. Statistical power depends upon how common is the occurrence of the outcome. Since the target complication rate in carotid endarterectomy is low (6%), monitoring outcome will only have limited ability to detect whether an individual surgeon’s true complication rate is greater than 6%. For example, if a surgeon’s ‘true’ complication rate was 8%, one would need to monitor the outcome of 1200 operations to detect that this surgeon’s rate was greater than 6% with 80% power at a significance level of 5%. Thus, in circumstances where the expected outcome is relatively rare, outcome indicators will only have limited power to detect real differences in quality.

Finally, one needs to consider whether there are any practical alternatives to using outcome indicators in a given area. If there is no evidence that a measure of process is linked to outcome, then there is little justification for using such a measure, unless perhaps if it is endorsed by a nationally credible consensus guideline. In such circumstances, outcome measures may be most appropriate.
Conclusion

Outcome measures are an indicator of health. They are valid as performance indicators in as much as the quality of health services has an impact on health [30]. In some circumstances, the quality of health services has a relatively minor role in determining health outcome, in other circumstances, a major role. Where health services have major effects on outcome, use of outcome measures as performance indicators is appropriate and efforts should be taken to ensure that the data can be interpreted reliably (e.g. standardization of data collection and development of robust case mix adjustment systems). Where factors such as lifestyle and socio-economic circumstance rather than health care are the major determinants of outcome, it would be a misnomer to refer to an outcome measure as a performance indicator, since it would be acting as a broader barometer of the health of that population [31]. This is not to say that health outcome data should not be obtained (and indeed published) in such circumstances. The purpose of collecting and publishing such data should not be to identify poorly performing hospitals, but rather to inform upon wider aspects of health policy. It would be inappropriate to develop complex case mix adjustment systems, since such adjustments will remove those very factors (e.g. smoking rates; socio-economic status) that ought to be under scrutiny.

Process measures are direct measures of the quality of health care, provided that a link has been demonstrated between a given process and outcome. Therefore, where such measures are available and they are relevant and practical, then they should be used in preference to outcome measures since they are much easier to interpret and are much more sensitive to differences in the quality of care.

While it is tempting to use mortality variations as a way of identifying poor performers, especially in the light of the Bristol case and the Shipman case in the UK, the appropriateness of such a strategy needs to be considered for each specific disease and operation. Alternative strategies such as process measurement or risk management techniques that focus on the analysis and investigation of individual incidents rather than seeking to interpret statistical variations may be more effective at protecting the public from poor quality care [32].

Acknowledgements

I would like to acknowledge the helpful comments of Richard McManus on drafts of this paper.

References

21. Marcinial TA, Ellerbeck EF, Radford MJ et al. Improving the

22. Mant J, Hicks NR, Fletcher J. Study should have had more patients or longer time scale. Br Med J 1996; 313: 1006.


Accepted for publication 3 July 2001