Quality indicators for general practice: which ones can general practitioners and health authority managers agree are important and how useful are they?

S. M. Campbell, M. O. Roland, J. A. Quayle, S. A. Buetow and P. G. Shekelle

Abstract

Background The aim of the study was to assess the face validity of quality indicators being proposed for use in general practice by health authorities.

Method A national survey of health authorities was carried out to identify quality indicators being proposed for use in general practice. A two-stage Delphi process was used to establish general practitioners' (GPs') and health authority managers' views on the face validity of identified indicators. A total of 240 separate indicators identified by health authorities and the NHS Executive as potential markers of the quality of general practice care were assessed. Indicators related to access, organizational performance, preventive care, care for a small number of chronic diseases, prescribing and gatekeeping. The subjects were a purposive sample of 47 health authority managers and 57 general practice course organizers.

Results Thirty-six indicators received median validity scores of 8 or 9 out of a maximum possible score of 9. Of this set, 83 per cent was rated identically by both groups of respondents. Prescribing and gatekeeping indicators generally received low validity scores.

Conclusion Acceptable face valid indicators were identified for all domains except gatekeeping. However, the indicators rated by the sample do not cover all aspects of care. No indicators were proposed for use by health authorities relating to effective communication, care of acute illness, health outcomes or patient evaluation. Although it is possible to develop indicators of general practice care which have face validity in the view of both GPs and managers, these will be very partial measures of quality. In the indicators used in this study, no explicit distinction was made between indicators designed to assess minimum standards with which all practices should comply, and indicators which could be used to reward higher levels of performance. Failure to separate these will result in antagonism from practitioners to quality improvement initiatives in the NHS, and a failure to engage the profession in improving quality of care.

Keywords: General practice, quality indicators, performance, Delphi

Introduction

Since the late 1980s quality improvement has become part of a government strategy for reforming the management and organization of public services, as part of a new approach to the public management of public services. Performance indicators were applied to hospitals during the late 1980s, with the development of indicators for acute hospitals. In 1996, health authorities in England and Wales were encouraged to develop indicators for assessing the quality of primary care, and this focus was given added impetus by the formation of an NHS Committee to develop a set of Primary Care Effectiveness and Efficiency indicators to be applied at the health authority level. The focus upon quality has been given further prominence with the White Paper The New NHS, which places quality 'at the heart' of the National Health Service, and the proposed National Performance Framework, which includes 81 'high level' indicators designed for use at the health authority level.

This paper reports the findings of a two-stage postal Delphi survey. It must be emphasized at the onset that the study did not seek to generate a set of indicators but rather to assess the face validity of actual indicators being used or planned for use by health authorities. The study had three aims:

1. The perspectives of health services managers and health
Table 1 Definitions of accessibility, clarity, reliability and validity

<table>
<thead>
<tr>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility was defined in terms of whether the information would be available currently or could potentially be collected easily</td>
</tr>
<tr>
<td>Clarity was defined as meaning the indicator was expressed in clear, precise and unambiguous language</td>
</tr>
<tr>
<td>A reliable indicator was defined as one for which the information required to apply the indicator was likely to be collected and applied in the same way by different users over time, and to be relatively error free</td>
</tr>
<tr>
<td>Validity was defined as meaning the indicator measured quality of general practice care as defined for the relevant domain listed in the Appendix</td>
</tr>
</tbody>
</table>

professionals have been found to diverge about the meaning of quality and approaches to its improvement. We sought to identify indicators commanding broad support from both health services managers and general practitioners (GPs) as indicators of quality of general practice care.

We sought to identify indicators commanding broad support from both health services managers and general practitioners (GPs) as indicators of quality of general practice care.

Quality indicators are often criticized as representing a partial perspective on quality. As such, we sought to highlight areas which are, and equally importantly, are not addressed by currently available indicators but are nevertheless important areas of quality of care.

The overall aim of the study was to generate a set of face valid indicators from this database of health authority derived indicators and place them in the public domain for discussion. Not all these indicators can be measured from currently available data: further work is needed to operationalize a number of indicators described in the paper.

Method

In August 1996, we wrote to all district primary care leads in England and Wales, asking for details of indicators which they were using or considering using in primary care. We received replies from 45 health authorities out of 106 contacted. To our knowledge, these respondents included all authorities active in the field of quality assessment in primary care. We telephoned a random sample of 10 non-responders: seven of these had not started developing indicators at the time of the survey. We believe that the dataset generated was therefore representative of the types of indicators being used or in the planning stage by health authorities. In consequence, the absence of indicators relating, for example, to the quality of the consultation was therefore a natural reflection of their absence in health authorities' sets of used or planned indicators. Moreover, most indicators identified were practice and population level indicators rather than consultation or individual patient specific indicators.

After removing duplicates, 240 separate indicators were identified. We also included a list of indicators proposed for use at health authority level by the NHSE Efficiency and Effectiveness Steering Group which had been circulated to health authorities for discussion, several of which have since been published in the National Performance Framework.

Definitions of high-quality care covering six domains was developed by a small expert advisory reference group (Appendix) and the indicators were allocated to these domains by the project team. These domains are comparable to four of the six areas stressed in the recent White Paper (Appendix).

We aimed to achieve 100 responses to the two rounds of the survey, approximately equally divided between GPs and health authority managers using a purposive sample. We sampled from 58 health authority managers having lead responsibility for primary care in their district and from a random sample of 89 general practice course organizers for a two-stage postal Delphi survey to assess the face validity of the proposed indicators. For health authority managers, we selected the 45 who had sent details of their authority's proposed quality indicators, and randomly selected an additional 13 (from a sample of 60). The health authority sample were given the opportunity to delegate responsibility for completing parts of the Delphi to others if they were thought to be better informed (e.g. pharmaceutical advisor completing the prescribing indicators section). General practice course organizers are responsible for vocational training for GPs in their district; we chose them as professional leaders who were likely to be familiar with concepts relating to quality measurement in general practice.

Table 2 Number of indicators and the proportion rated face valid (given median scores of 8 or 9) in each domain

<table>
<thead>
<tr>
<th>Domain of quality</th>
<th>Number of indicators</th>
<th>% of indicators with median score 8 or 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>29</td>
<td>35</td>
</tr>
<tr>
<td>Organizational performance</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td>Preventive care</td>
<td>13</td>
<td>39</td>
</tr>
<tr>
<td>Chronic disease management</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>Prescribing</td>
<td>45</td>
<td>11</td>
</tr>
<tr>
<td>Gatekeeping</td>
<td>24</td>
<td>0</td>
</tr>
</tbody>
</table>
In the first round, respondents were asked to rate the validity and clarity of each indicator against the definitions in Table 1. Each attribute was measured on a nine-point scale. After the first round, we removed 62 indicators which failed to meet a pre-set median score of six for validity, and where the respondents had indicated that the meaning of indicators was unclear, we clarified or reformulated some of the remainder. Indicators which were statutory requirements were also excluded as they were not likely to be discriminating measures of quality.

Panellists were given feedback (using median scores) on their validity scores in round 1, and asked to re-rate the 201 indicators in round 2 for validity, reliability and accessibility, again on separate nine-point scales using the definitions in Table 1.

A total of 57 general practice course organizers (64 per cent) and 47 health authority managers (81 per cent) completed the second round of the survey. Scores were analysed using SAS, with Mann–Whitney U-tests being used to test for significant differences between the median scores of GPs and health authority managers. In this paper, we present the data on the indicators which were given highest ratings of face validity by our panel in the final round (median score 8 or 9 out of a maximum of 9).

Results

We defined an indicator as face valid if it was rated with a median score of 8 or 9. The number of indicators and the proportion identified as face valid in each domain during Round 2 are shown in Table 2. Table 3 shows the 36 indicators rated as face valid. These are discussed for each domain with additional results. A full list of scores for all 201 indicators is available on the NPCRDC web-site (http://www.npcrdc.man.ac.uk) or from the authors.
Table 3 continued

<table>
<thead>
<tr>
<th>Items receiving scores of 8 or 9 for validity</th>
<th>Validity median score</th>
<th>Overall median score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GPs</td>
<td>Managers</td>
</tr>
<tr>
<td></td>
<td>Validity</td>
<td>Reliability</td>
</tr>
<tr>
<td>Chronic disease management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice maintains a register of asthma patients</td>
<td>8 (8–8)</td>
<td>8 (8–8)</td>
</tr>
<tr>
<td>≥70% asthma patients reviewed in last 12 months</td>
<td>8 (7–8)</td>
<td>8 (8–8)</td>
</tr>
<tr>
<td>Practice should have a nebulizer or nebulhaler, or volumatic and oral steroids available</td>
<td>9 (9–9)</td>
<td>8 (7–8)</td>
</tr>
<tr>
<td>Practice has a stepped care approach asthma protocol based on BTS guidelines</td>
<td>8 (8–8)</td>
<td>8 (8–8)</td>
</tr>
<tr>
<td>Practice maintains a register of patients with diabetes</td>
<td>8 (8–9)</td>
<td>8 (8–9)</td>
</tr>
<tr>
<td>≥80% diabetic patients (insulin dependent and NIDDM) reviewed in past year</td>
<td>8 (8–8)</td>
<td>8 (8–8)</td>
</tr>
<tr>
<td>All diabetic patients should receive an annual review covering: glycaemic control, eyes, blood pressure, feet, renal function, and cardiovascular risk factors</td>
<td>8 (8–9)</td>
<td>8 (8–9)</td>
</tr>
<tr>
<td>Adequate patient information should be available in a suitable format for patient self-management of diabetes</td>
<td>8 (8–8)</td>
<td>8 (7–8)</td>
</tr>
<tr>
<td>Practice maintains a register of hypertension patients</td>
<td>8 (7–8)</td>
<td>8 (7–8)</td>
</tr>
<tr>
<td>≥80% patients with hypertension reviewed in the last 12 months</td>
<td>8 (7–8)</td>
<td>8 (7–8)</td>
</tr>
<tr>
<td>Effective hypertension management – annual calibration of sphygmomanometers</td>
<td>8 (8–8)</td>
<td>7 (7–8)</td>
</tr>
<tr>
<td>Practice has a register of the age–sex distribution of the practice population</td>
<td>8 (8–9)</td>
<td>8 (7–9)</td>
</tr>
<tr>
<td>Prescribing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice carries out audit of repeat prescribing</td>
<td>8 (7–8)</td>
<td>8 (8–8)</td>
</tr>
<tr>
<td>Practice uses PACT data to help review its prescribing behaviour</td>
<td>8 (8–8)</td>
<td>8 (8–8)</td>
</tr>
<tr>
<td>Ratio of co-trimoxazole to trimethoprim</td>
<td>8 (7–8)</td>
<td>8 (7–8)</td>
</tr>
</tbody>
</table>

Health authority managers rated 26 per cent of the indicators face valid compared with 21 per cent for general practice course organizers. Health authority managers rated indicators higher than GPs overall (median scores: health authority managers 6.74, general practice course organizers 6.22; p < 0.0001). Where differences in ratings occurred for the complete set of 201 indicators in the final round, in four-fifths of cases the health authority managers’ scores were higher. This suggests that GPs were more doubtful about the validity of these indicators. However, 28 of the 36 face valid indicators in Table 3 were given the same median score by GPs and health authority managers, and the remainder had scores which differed by one point or less; suggesting that the set of highest rated indicators listed in Table 3 would command broad support as indicators of quality for general practice.

Access

There was relatively good agreement between GPs and health authority managers for the indicators relating to access. Both groups gave high scores to access to the practice by telephone throughout the working day, and to the availability of telephone advice for patients. In terms of consulting sessions, both believed that surgeries sessions should be available for at least four mornings and four afternoons, and that routine appointments should be of at least 10 minutes duration. Median ratings of 8 were given for disabled access and 7 for access to both male and female GPs, though both these are probably more appropriately applied to localities rather than individual practices. Median ratings of 7 were recorded for patients not having to wait more than 20 minutes for their appointment and for routine appointments being available within three working days. Adequate privacy in the reception area scored a median rating of 8 but we have not included it in Table 3 because of perceived difficulty in measuring it.

Organizational structure

Indicators relating to organizational structure received only moderate support as indicators of quality, with only 16 per cent receiving scores of 8 or 9. A substantial number of indicators in this domain related to the number of GPs or employed staff (either per 1000 population or per GP): few of these scored highly for face validity.

Higher scores were given to indicators relating to training (e.g. appropriate skills, audit programme, commitment to
staff training) and clinical audit, though these scored lower in terms of the accessibility or reliability of currently available information.

Organizational structure: range of services
On site family planning services were rated highly, but on site provision of other individual services was given lower ratings (minor surgery 7, physiotherapy 6, counselling 6, chiropody 6). However, the importance of providing access to a wide range of services was emphasized by the score of 8 for 'practice registered for child health surveillance, minor surgery and maternity care', which was one of the few indicators where the GPs gave higher ratings than the health authority managers.

Preventive care
The highest validity rating for indicators by domain occurred in this category, with 39 per cent of indicators receiving scores of 8 or 9. Many of the indicators in this area not included in Table 3 also related to immunization and cervical cytology screening. Indicators related to hypertension and obesity screening were rated with medians of 7 and 5.5, respectively.

Chronic disease management
Thirty-five per cent of indicators relating to chronic disease management had median scores of 8 or 9. These indicators related substantially to organization of care. For example, disease registers and evidence of regular follow-up were common themes for high scoring items for diabetes, asthma and hypertension. These were rated higher than indicators which related, for example, to individual aspects of prescribing.

Prescribing
There was a relatively low level of support for prescribing indicators. The higher scoring items related again to the way in which care was organized in the practice (e.g. prescribing audit, repeat prescribing system in operation), rather than to individual prescribing indicators. The use of PACT data has been found by others to be a useful mechanism for audit.18 Of prescribing indicators which scored 7, a substantial number again related to organization and training; for example, regular prescribing meetings, practice formulary, guidelines for locums. An indicator concerning the existence of a repeat prescribing system in practices scored a median of 8 but was not included in Table 3 because of difficulty in meaningful implementation.

Only one indicator (Table 3) relating to individual drug details scored 8. Others which scored 7 included the percentage of generic prescribing, lipid lowering drugs per ASTRo-Pu, benzodiazepines per ASTRo-Pu, bronchodilator:steroid/cromoglycate ratio, and 'plain':combination diuretic ratio. The other 24 indicators relating to data which could be derived from routine PACT analysis were all given lower scores.

Gatekeeping
No indicators relating to gatekeeping were given high scores for face validity by the panel. The indicators rated in this domain received the least support, with 92 per cent of indicators rated at 6 or below. These in general related to hospital referral or admission rates: several of these have subsequently been published in the National Performance Framework consultation document.15 Admission rates for hip replacement, cataracts or insertion of grommets all scored 5, as did practice referral rates. The low scores for these items probably reflect the poor relationship between rates of referral and quality of care19,20 and the difficulty of applying these at practice level, and the fact that they are likely to be influenced by population and secondary care characteristics. Indicators of this sort are more appropriately regarded as indicators of the whole health care system and not just primary care.

Discussion
The legitimacy and utility of quality indicators depends on their acceptability to the various constituencies which they will affect. We believe that if indicators can be developed which are both valid and acceptable to both assessed and assessors, there is more chance of the findings being used to facilitate quality improvement. Because GPs12 and health authority managers21 may be suspicious about the prospect of quality indicators being used over which they have no ownership, and because of differences in their perceptions of quality,13 we sought to identify indicators which would command broad support. We believe that the set of indicators listed in Table 3 can be seen as face valid indicators of quality for general practice, judged by their high scores for validity. However, we do not claim that this set represents a comprehensive assessment of quality of care in general practice. Rather, there are several reasons to view this set with caution.

The set of indicators rated valid may be criticized for being unrepresentative. First, the sample of managers and GPs was purposive and it may be that another sample may have rated a different set of indicators as valid. However, we deliberately selected only the highest scoring indicators, as these are known to show greatest reproducibility between different panels.22 Second, other sources such as Royal Colleges or CME units could have been surveyed to identify indicators. However, the aim of the research was to rate indicators being used or in development by health authorities and to highlight areas not covered. The indicators rated in this study clearly do not represent the full spectrum of general practice care. For example, within structural aspects of general practice, access, standards of premises and staff characteristics were all included to some degree. In terms of the process of care there were a
substantial number of indicators relating to preventive care and chronic disease management. However, the indicators being considered for use by health authorities contained some notable absences that are essential components of quality of care. For example, no indicators were rated relating to interpersonal skills, the quality of the actual consultation, the interface with social care and social services; or two areas of the ‘New NHS Performance Framework’ – outcome (either in terms of health status or patient evaluation) and the patient–carer experience. The absence of indicators relating to acute care management, the very essence of general practice, or indicators relating to good communication was also notable. We believe that aiming to assess the overall quality of care in general practice requires the use of broader sets of indicators. Population-based standardized measures such as those in this set may have an important role to play. However, they fail to reflect the complex needs of individual patients and the diversity of local circumstances, for which other sets of indicators are required.

Although such absences relate to the way in which our sample was drawn, the difficulty of developing indicators which cover more than a proportion of issues related to quality has been found elsewhere. For example, in the United States, the HEDIS 3.0 set of indicators, which has been extensively researched and is now being widely applied, relates almost exclusively to access and preventive care. To believe that a great deal more work is necessary before it will be possible to obtain more than a very partial view of the quality of general practice care.

We defined an indicator using the EQuiP Working Party definition: ‘a measurable element of practice performance for which there is evidence or consensus that it can be used to assess the quality, and hence change the quality, of care provided’. To be usable, indicators must relate to a clearly defined aspect of care, be sensitive to and discriminate between practices, and be easily quantifiable. Some of the indicators in Table 3, although rated valid, do not satisfy these characteristics. Some may be seen as activity rather than quality indicators. For example, those relating to chronic disease registers show the presence or absence of such features but do not give insight into the accuracy of the register or how it is used in practice. Moreover, for indicators to be meaningful they must be measurable. In general, the indicators identified in Table 3 are not available nationally as routine data and several would be very difficult to collect, even though they relate to important aspects of care (e.g. staff possessing appropriate skills). New types of data will need to be collected to deliver reliable and valid assessments of quality of general practice care, rather than relying, as in the past, on introducing indicators because the data are available. As the White Paper emphasizes, there should not be a ‘narrow obsession with counting activity for the sake of it’ and this will require the development of new measures.

Furthermore, indicators may be influenced by factors outside the control of individual practices. Environmental factors such as the socio-demographic characteristics of the population are known to have a major effect on, for example, cervical cytology rates, immunization and screening. Improvement is a relative as well as an absolute concept. These factors must be built into any quality assessments. Quality indicators do not define the cause of a problem: they identify an issue (which may not be the same in different practices) which may require further investigation. Problems may lie with the practice (e.g. in relation to prescribing patterns), with the health authority (e.g. in relation to investment in premises), or outside the complete control of either (e.g. in relation to screening uptake).

It is also important to emphasize that indicators may be used legitimately by health authorities for different purposes; for example, contract monitoring, monitoring quality of access in areas of under-developed general practice such as in inner cities, or using sets of indicators as conduits for quality improvement. What is crucial is that the purpose to which indicators will be put is stated explicitly and that indicators are validated specifically for each of these different purposes. Moreover, findings from using indicators are the beginning and not the end of a process of enquiry.

In the indicators used in this study, no explicit distinction was made between indicators designed to assess minimum standards (with which all practices should comply), and indicators designed to assess higher levels of performance. We believe that this is a crucial distinction. Performance management of the NHS and the introduction of clinical governance require two explicit and separate approaches to quality measurement. Minimum criteria should be set low, and should be readily achieved by the great majority of practitioners. High-quality performance should be separately and explicitly rewarded. Failure to separate these will result in antagonism from practitioners to quality improvement initiatives in the NHS, and a failure to engage the profession in improving quality of care.

The results reported in this paper show that it is possible to start to develop measures of quality of general practice care which are perceived to be valid by, and acceptable to, both health authority managers and GPs for a limited number of domains of quality. However, we believe that the proposed set demonstrates the importance and necessity of using a variety of different approaches to quality measurement and to be explicit about what is and what is not being measured. We hope that the indicators listed in Table 3 can contribute to a continuing debate.

Acknowledgements

We are very grateful to the GPs and health authority managers who completed the lengthy questionnaires in this study. We are also grateful to our expert panel, both for their continued advice and for their contribution to the definitions in the Appendix:
they were Hilary Hearnshaw, Martin Lawrence, Geoff Meads, Jonathan Shapiro, George Taylor and Barry Thomas. Dr Shekelle is a Senior Research Associate of the US Department of Veterans Affairs Health Services Research and Development Service, and participated in this project during his tenure as an Atlantic Fellow in Public Policy, an international fellowship supported by the Foreign and Commonwealth Office.

References


Accepted on 2 June 1998

Appendix: Quality domains represented within indicator set and definitions

Terms in italics relate to areas of the White Paper Performance Framework.  

**Access: Fair access**

High-quality general medical services are physically accessible, available and meet demands in ways which are appropriate to need and local circumstances.

**Organizational performance: Efficiency, and effective delivery of appropriate health care**

High-quality general medical services demonstrate internal processes for planning, providing and reviewing care that deploy available resources appropriately and efficiently to meet local needs.
Service performance: preventive care: *Effective delivery of appropriate health care*

High-quality general medical preventive care informs, recommends and provides for all appropriate people an integrated and appropriate package of care that can prevent the onset or progression of disease while respecting the rights of patients to make their own choices.

Service performance: chronic disease management: *Effective delivery of appropriate health care*

Within an integrated, comprehensive and appropriate package of care, high-quality general medical services address the physical, psychological and social needs of patients with chronic illness and their carers.

Prescribing: *Effective delivery of appropriate health care, and efficiency*

In high-quality general medical services prescribing decisions can be demonstrated to maximize effectiveness, safety and economy, and to respect patients’ informed choices.

Gatekeeping: *Effective delivery of appropriate health care*

High-quality referral decisions from general medical care providers should be appropriate for the particular patient, and be timely in relation to the course of the condition. The potential benefits of the referral should justify the likely costs.