

# Assessing the quality of care of multiple conditions in general practice: practical and methodological problems

S A Kirk, S M Campbell, S Kennell-Webb, D Reeves, M O Roland, M N Marshall

*Qual Saf Health Care* 2003;12:421–427

See end of article for authors' affiliations

Correspondence to:  
Dr S A Kirk, School of Nursing Midwifery and Health Visiting, Coupland 3, University of Manchester, Oxford Road, Manchester M13 9PL, UK; sue.kirk@man.ac.uk

Accepted for publication 4 June 2003

**Objective:** To investigate practical and methodological problems in assessing the quality of care of multiple conditions in general practice.

**Setting:** Sixteen general practices from two socioeconomically diverse regions in the UK.

**Method:** Quality of care was assessed in 100 randomly selected patient records in each practice using an established set of quality indicators covering 23 conditions commonly seen in primary care. Inter-rater reliability assessment was carried out for five of the conditions.

**Results:** Conducting simultaneous quality assessment across multiple conditions is highly resource intensive. Poor data quality and the low prevalence of some items of care defined by the indicators are significant problems. Scores for individual indicators require very large samples for reliable assessment. Quality scores are more reliable when reported at a higher unit of analysis. This is particularly true for indicators and conditions with low prevalence where data may need to be aggregated to the level of groups of conditions or organisational providers. There is no single ideal way of aggregating quality scores.

**Conclusion:** The study identified some of the practical and methodological difficulties in assessing quality of care across multiple conditions. For improved quality assessment, advances in information technology and improvements in data quality are required for more efficient and reliable data extraction from medical records, together with the development of methods for combining scores across indicators, conditions, and practices. However, electronic data extraction methods will still be based on the assumption that the care recorded reflects the care provided.

Quality assessment plays an important part in any attempt to improve the quality and increase the accountability of healthcare organisations.<sup>1,2</sup> In recent years rapid advances have been made, particularly in relation to developing quality indicators for condition specific clinical processes.<sup>3–5</sup> While these indicators have helped focus attention and improve our understanding of quality improvement in primary care, there are still many challenges that need to be addressed.

One challenge is a conceptual one: quality assessment that focuses on single diseases or conditions in isolation fails to reflect the nature of primary care.<sup>6</sup> Patients presenting to primary care practitioners usually have more than one clinical condition.<sup>7,8</sup> Co-morbidity is particularly important in an aging population and single disease quality assessment fails to recognise this complexity.<sup>9</sup>

There are also methodological challenges. The development of quality indicators, even using the most rigorous approaches available, represents only the first stage in the process of quality assessment.<sup>10</sup> If indicators can be developed for a range of conditions, is it feasible to extract the required data from the medical record and is this extraction process reliable? Should indicators be applied to all patients with the given condition, or to a sample of those patients, and how should such samples be selected? How can results across a range of indicators or conditions be summarised?

There are both practical and policy related reasons for addressing these issues. Primary care professionals tend to criticise attempts to package their work into single biomedical conditions and this is one of the reasons why they may not engage with current quality assessment initiatives.<sup>11,12</sup> This problem has resulted in calls from professionals for simultaneous assessment across multiple conditions.<sup>9</sup> At the same time, policy makers are becoming aware that single

condition assessment can encourage practitioners to focus their attention on the condition that is being measured, to the detriment of conditions that are not being assessed.<sup>13,14</sup>

An increasing range of methods is being developed in order to make external judgements about the quality of the care being provided. This demands greater rigour than has been shown in the past, both in terms of conceptualising the role of measures and in understanding the science underlying them.

In this paper we investigate the practical and methodological problems of conducting a multi-condition quality assessment in UK general practice using a published set of clinical quality indicators.<sup>5</sup>

## METHODS

### The quality indicator set

We conducted the quality assessment exercise using a published set of quality indicators representing 23 of the most common conditions seen in general practice.<sup>5</sup> The conditions cover acute, chronic and preventative care (table 1) and we estimate from routine data sources<sup>7</sup> that they address approximately 65% of the problems presenting to general practice. The indicators were developed using a modification of the RAND appropriateness method,<sup>10,15</sup> an approach which ensures that the resulting indicators have high face validity among health professionals. This process is described in detail elsewhere.<sup>3,5</sup> Indicators were only accepted if there was professional consensus that each represented an important element of quality for that condition, that the relevant information needed to be recorded in the patient record, and that the indicator was supported by the best available scientific evidence. For this reason, where the evidence is good and professional consensus strong—for example, diabetes mellitus (box 1) or

**Table 1** Results of quality assessment

	No (%) of patients with condition	No of indicators in condition set	No of patients with at least one applicable indicator	No of indicators that applied to ≤10% of patients with condition	No of indicators that applied to >50% of patients with condition	Composite pass rate (%)*	Range of pass rates for individual indicators (%)
<b>Chronic conditions</b>							
Acne	25 (1.6%)	5	7	3	–	25	0–43
Allergic rhinitis	36 (2.3%)	4	2	4	–	100	100–100
Asthma	194 (12.1%)	24	194	10	8	61	0–100
Coronary artery disease	53 (3.3%)	13	53	3	7	76	29–91
Depression	141 (8.8%)	21	141	12	5	52	14–100
Diabetes	34 (2.1%)	14	34	2	9	82	47–100
Dyspepsia/PUD	138 (8.6%)	7	138	1	3	24	9–100
Hypertension	149 (9.3%)	26	149	9	4	69	0–100
Osteoarthritis	91 (5.7%)	4	72	1	1	45	27–100
Total (chronic)	861 (53.8%)	118	790	45	37		
<b>Acute conditions</b>							
Acute childhood diarrhoea	59 (3.7%)	8	59	1	4	32	6–50
Acute low back pain	212 (13.3%)	3	100	1	–	32	27–60
Headache	232 (14.5%)	7	13	7	–	93	67–100
Otitis media	68 (4.3%)	2	68	–	1	87	44–97
URTI - bronchitis	63 (3.9%)	3	12	1	–	35	0–71
URTI - rhinitis	3 (0.2%)	1	2	–	1	100	100–100
URTI - tonsillitis	142 (8.9%)	3	141	1	1	79	73–100
UTI	195 (12.2%)	20	148	17	–	53	0–100
Total (acute)	974 (60.9%)	47	543	28	7		
<b>Preventive care</b>							
Cervical screening	429 (26.8%)	6	355	4	1	91	7–100
Family planning	90 (5.6%)	5	89	2	3	71	0–83
Influenza immunisation	324 (20.3%)	7	323	3	1	56	35–100
HRT	52 (3.3%)	5	51	1	3	54	0–71
Pneumococcal immunisation	259 (16.3%)	7	256	4	1	26	19–100
Hepatitis B immunisation	5 (0.3%)	5	5	4	1	20	0–25
Total (preventive)	1159 (72.4%)	35	1079	17	10		

URTI=upper respiratory tract infections; UTI=urinary tract infections; HRT=hormone replacement therapy.

\*Total number of indicators met, summed across all indicators and patients and divided by total number of indicators that applied (expressed as a percentage).

coronary heart disease (CHD)—the indicator set covers most of the elements of high quality care for that condition. In contrast, for conditions with a weak evidence base and poor consensus about the elements of high quality care—for example, acne vulgaris (box 2) or headache—the indicator set is small and addresses only a small part of the care for the condition.

### Practice selection

The audit was conducted in two Primary Care Trusts (PCTs) (see box 3), one in a deprived inner city area in north-west England and the other in an affluent semi-rural area in south-west England. The local research ethics committees in each locality granted approval to conduct the study. A random sample of 10 practices in each PCT, stratified by practice size and teaching status, was invited to participate. Sixteen of the practices (80%, nine in one PCT and seven in the other) agreed to take part. One of the practices had no computer, nine used their computer to record all patient contacts, and six used it only for registration and repeat prescribing.

### Patient record sampling

We selected a random sample of patient records from each practice, an approach that has been used in other large scale quality assessment exercises.<sup>17–18</sup> One hundred records (both electronic and paper) were sampled from each practice, giving 1600 records in total. We estimated that, given the prevalence of the chosen conditions, this number of records would ensure that all of the conditions had a high chance of being sampled in each of the practices. Random samples were identified either using the practice's own computer system or by using random number tables in order to select from lists of patients where an electronic facility was not available.

### Data collection

Data were collected by two researchers (SK, SKW) between October 2000 and May 2001, one working in each locality. Both had a background in nursing and both were trained to review patient records. For each patient, both the electronic and paper records were examined to determine: (a) whether there were any entries in the previous 5 years relating to the tracer conditions; (b) whether the indicators for identified conditions could be applied; and (c) whether or not the indicators were met. Data were anonymised before being removed from the practice. Inter-rater reliability was formally assessed for five of the conditions (depression, hypertension, upper respiratory tract infections, family planning, and urinary tract infections). To do this, the two researchers independently extracted data from 25 sets of records in each of four practices and their results were compared.

### Data analysis

The data were analysed using SPSS version 10.1 for basic frequency counts and STATA version 8 for statistical modelling. Inter-rater reliability was assessed using the Cohen kappa coefficient of agreement.

The aim of this study was to investigate the feasibility, in terms of both practical and methodological problems, of conducting a multi-condition quality assessment in UK general practice using a published set of clinical quality indicators.<sup>5</sup> Using the data collected in the study we explore the feasibility of record sampling, the reliability of data extraction, and the methodological difficulties inherent in calculating summary scores. In addition, we use the data to examine the number of patient records that need to be reviewed in order to produce statistically reliable quantitative measures of quality and the issue of data quality within records.

**Box 1 Indicator set for diabetes mellitus: example of a condition with a good evidence base and strong consensus about the elements of high quality care**

- The diagnosis of diabetes should be clearly identifiable on the electronic or paper records of all known diabetics.
- If the HbA1c level of a diabetic patient is measured as >8%, the following options should be offered: change in dietary or drug management; explanation for raised test; or written record that higher target level is acceptable.
- HbA1C levels should be checked in diabetic patients at least every 12 months.
- If a diabetic has a sustained blood pressure recorded as >140/85 mm Hg on three or more consecutive occasions, then a change in non-drug or drug management should be offered.
- Diabetics should have their feet examined at least once every 12 months.
- If there is evidence of foot deformities, history of foot ulceration, significant vascular or neuropathic disease, the patient should be referred to an appropriate service if not already under their care.
- All diabetic patients should have an annual fundal examination.
- All diabetic patients should have the following measurements taken for lipid profile within the last 3 years: total serum (1) cholesterol (2) triglycerides
- Diabetic patients with established ischaemic heart disease and a raised fasting cholesterol ( $\geq 5$  mmol/l) should be advised about dietary modification or to take lipid lowering medication.
- Diabetic patients with sustained proteinuria should be currently prescribed treatment with ACE inhibitors unless contraindicated.
- Patients should be seen by an appropriate health care professional (GP, practice nurse, diabetic doctor) annually.
- All diabetic patients should be offered influenza vaccination annually and pneumococcal vaccination unless contraindicated or intolerant.

## RESULTS

The feasibility of multi-condition quality assessment is examined in relation to four issues:

- Accessing patient records and the quality of data within them.
- The reliability of data extraction.
- Varied prevalence rates for conditions and indicators.
- Construction of composite quality scores.

### Access to and quality of patient records

Records for all sampled patients could be identified. Manual data extraction was labour intensive and highly dependent on the quality of the records and the number of conditions in each record. The practices were all supportive of the study but lack of space or access to a computer terminal often meant that the researcher had to be sensitive about the demands that they were making on the practices. On average, it took

**Box 2 Indicator set for acne vulgaris: example of a condition with a weak evidence base and poor consensus about the elements of high quality care**

- Oral tetracycline should not be prescribed for adolescents under 12 years of age.
- If oral tetracycline is prescribed for a female of childbearing age (16–45 years), enquiry should be made about the date of last menstrual period or a negative pregnancy test.
- If oral tetracycline is prescribed for a female of childbearing age (16–45 years), advice should be given regarding effective means of contraception (including abstinence).
- If topical retinoids are prescribed to females of childbearing age (16–45 years), enquiry should be made about the date of last menstrual period or a negative pregnancy test.
- If topical retinoids are prescribed to females of childbearing age (16–45 years), advice should be given regarding effective means of contraception (including abstinence).

5–10 days to audit 100 records in each practice. The length of time to examine each set of records ranged from a few minutes to over an hour. Even in the practices that made significant use of their computer, some data items could only be found in the manual records.

### Reliability of data extraction

Inter-rater reliability was assessed for five conditions which included 52 indicators. Of these, 10 indicators had a kappa of less than 0.6 indicating poor reliability of data extraction. Six of these related to depression. These 10 indicators were excluded from the subsequent analysis and are not presented in this paper.

### Condition and indicator prevalence

As expected, the prevalence of the 23 conditions varied considerably although the overall prevalence rates across the 16 practices (table 1) were in line with published data.<sup>7</sup> Not surprisingly, indicators relating to preventative care, which

**Box 3 Roles and responsibilities of Primary Care Trusts**

Primary Care Trusts are National Health Service organisations that have the main responsibility for assessing need, planning, and securing of all health services and improving health in England. They aim to actively engage with local communities and professionals and lead the NHS contribution to joint work with local government and other partners. Their functions are:

- Improving the health of the local community.
- Securing the provision of services.
- Integrating health and social care in the local health and social care community.

They are accountable to the Secretary of State for Health through regionally based strategic health authorities.<sup>16</sup>

apply to a large proportion of the population—for example, immunisation or cervical screening—could be applied most frequently. Where the indicator set for a given condition referred to only a small part of the care for that condition, the overall prevalence for that condition appeared to be low even though the condition was common. For example, the indicators for “headache” referred only to the management of head injury and migraine, so although 232 of the 1600 records had a reference to headache, the indicator set could be applied to only 13 of these records. In contrast, conditions with a relatively low prevalence but with a comprehensive set of indicators that could be applied to most or all of the records had a much higher profile. For example, only 34 of the 1600 records examined triggered the diabetes indicator set but the 13 indicators could be applied 344 times to these records.

### Quality scores

For any individual indicator the most intuitive and simplest way of constructing a “quality score” is to compute the percentage of times that the indicator, when it applied, was met. This is referred to as the indicator “pass rate”. Thus, of 34 patients with diabetes mellitus who should have had an annual fundal examination, 29 had evidence of one in the medical records, yielding a pass rate of 85%. For any one condition a wide range in the pass rates for individual indicators was observed (table 1). The last but one column in table 1 gives a “composite” pass rate for each condition, computed as the total number of passes across all indicators for that condition divided by the total number of times those indicators applied. The composite pass rates vary widely between conditions.

### Adequacy of sample size

The low prevalence of some indicators meant that some of the pass rates could not be estimated accurately. The composite pass rates, being based on pooled indicators, are more reliable than those for individual indicators. Using the study data it is possible to explore the sample size required to estimate reliable pass rates for practices. The question is important as it may not be feasible to assess the quality of particular aspects of care if the number of patient records that need to be sampled is very large.

Analyses are presented for individual indicators (table 2) and for composite “condition level” pass rates (table 3). This analysis can inform the use of quality scores as a “screen” to identify where quality of care in a practice falls below a certain minimum standard. With this in mind, we calculated sample sizes that would estimate indicator pass rates within a range of  $\pm 10\%$  with 90% confidence. So, if the minimum standard for a particular indicator was set at 60%, the data would identify 19 practices out of every 20 that were genuinely operating at a level 10 points below this minimum (that is, around 50%).

### Individual indicator level scores

To examine the pass rates expected for an “average” practice, the mean pass rate on each indicator was computed by pooling the data across all practices in the study. Individual indicators are binary variables (pass or fail) so it is reasonable to assume that the pass rate for an individual practice, based on a random sample of patient records, will have a binomial distribution. Using this assumption, the sample sizes required to achieve a 90% confidence interval of  $\pm 10$  points around the pass rate for each indicator are summarised in table 2.

**Table 2** Samples sizes required for individual level scores to achieve 90% confidence intervals of  $\pm 10$  points on estimated pass rates

	No of cases with the condition required to calculate reliable indicator level scores		
	No of indicators requiring a sample of $\leq 50$ cases	No of indicators requiring a sample of 51–100 cases	No of indicators requiring a sample of $>100$ cases
<b>Chronic conditions</b>			
Acne	–	–	5
Allergic rhinitis	–	–	4
Asthma	4	6	14
Coronary artery disease	5	2	6
Depression	3	2	16
Diabetes	7	2	5
Dyspepsia/PUD	2	2	3
Hypertension	2	2	22
Osteoarthritis	–	2	2
Total	23	18	77
<b>Acute conditions</b>			
Acute childhood diarrhoea	1	4	3
Acute low back pain	–	–	3
Headache	–	–	7
Otitis media	1	–	1
URTI – bronchitis	–	–	3
URTI – rhinitis	–	1	–
URTI – tonsillitis	1	1	1
UTI	–	–	20
Total	3	6	38
<b>Preventive care</b>			
Cervical screening	1	–	5
Family planning	1	1	3
Influenza immunisation	–	–	7
HRT	–	3	2
Pneumococcal immunisation	–	1	6
Hepatitis B immunisation	–	1	4
Total	2	6	27

URTI = upper respiratory tract infections; UTI = urinary tract infections; HRT = hormone replacement therapy.

The table shows that, for four of the 24 asthma indicators, the records of 50 asthma patients or fewer are required in order to estimate the true pass rate to the desired degree of accuracy; but for 14 indicators more than 100 sets of records are needed. Across all 23 conditions, 142 indicators out of 200 (71%) require a sample of more than 100 and 31% require more than 1000 records. The smaller the subset of patients to whom the indicator applies, the larger the number of records that have to be assessed in order to obtain enough examples where the indicator applies.

### Condition level scores

The composite pass rate for each condition (table 1) was taken to represent the score expected at the "average" practice. The sample size required to achieve a 90% confidence interval of  $\pm 10$  points around each score was then calculated. To do this we used the fact that the composite score for a condition can be expressed as a weighted mean of the pass rates for the individual patients with that condition (the weights being the number of indicators applying to each patient). The generalised linear modelling (GLM) procedure within STATA was used to model the proportion of indicators passed by each patient using the number that applied as weights, and an estimate obtained of the standard deviation of the weighted mean score. Although the patient level scores have an unknown distribution, under the central limit theorem their weighted mean will tend to a normal distribution as the sample size increases, and hence confidence intervals and sample sizes have been derived using an assumption of normality. Differences between practice means have been controlled for in the analyses, except where this left fewer than 30 degrees of freedom for estimating the standard deviation. Four conditions were represented by fewer than eight patients, too few to derive a reliable measure of standard deviation. To err on the side of caution, sample size requirements for these conditions have been estimated using

the 75th percentile of the standard deviations for the remaining conditions within the same modality.

The results are shown in table 3. It can be seen that the numbers of patients required are dramatically reduced in comparison with the numbers required for the individual indicators. For example, only eight cases are required for the assessment of asthma or coronary heart disease. Indeed, 10 of the 23 conditions require a sample of 20 cases or fewer to achieve the desired accuracy. In only four instances are more than 100 cases required, and these are all conditions where a high proportion of the indicators applied to only a subset of patients. Table 3 also shows the numbers needed if sampling could be restricted solely to patients with applicable indicators. In this situation, 13 conditions can be assessed using a sample of 20 and only five conditions require more than 40 cases. The penultimate column in table 3 shows the number of records that would have to be drawn at random to identify sufficient cases with each of the conditions to give a reliable overall score for that condition. All conditions except asthma and cervical screening require a sample of more than 100 randomly selected records. However, a single random sample of 500 would be sufficient to allow reliable quality scores to be computed for 14 of the 23 conditions.

### PCT level quality scores

The sample size estimates need to be increased when the objective is to derive quality scores for a group of practices (for example, a PCT) rather than a single practice. In this situation, cluster effects have to be taken into account and this can be done with the study data by examining the standard deviation for each condition level quality score. This was done separately for each of the two PCTs in the study and the results pooled to obtain a single estimate. The ratio of the sample sizes required at PCT level in comparison with practice level are shown in the right hand column of table 3. For most conditions the PCT samples need to be between one and two times the size of the sample required for individual practice level scores. For pneumococcal immunisation, the

**Table 3** Samples sizes required for condition level scores to achieve 90% confidence intervals of  $\pm 10$  points on estimated pass rates

	No of cases with the condition required	No of cases with applicable indicators required	No of randomly selected cases required	Multiplication factor for PCT level scores*
<b>Chronic conditions</b>				
Acne	60	17	3685	†
Allergic rhinitis	304	17	13526	†
Asthma	8	8	64	2.3
Coronary artery disease	8	8	233	1.3
Depression	14	14	155	1.3
Diabetes	7	7	342	†
Dyspepsia/PUD	17	17	196	1.0
Hypertension	12	12	127	2.4
Osteoarthritis	61	48	1071	1.0
<b>Acute conditions</b>				
Acute childhood diarrhoea	19	19	511	1.4
Acute low back pain	103	48	774	1.6
Headache	347	19	2391	†
Otitis media	14	14	334	1.6
URTI – bronchitis	217	41	5514	†
URTI – rhinitis	62	41	33085	†
URTI – tonsillitis	33	33	370	1.5
UTI	51	38	416	2.1
<b>Preventive care</b>				
Cervical screening	11	9	40	1.6
Family planning	20	20	357	1.9
Influenza immunisation	62	62	307	2.9
HRT	34	33	1052	1.0
Pneumococcal immunisation	37	37	231	8.3
Hepatitis B immunisation	37	37	11851	†

URTI = upper respiratory tract infections; UTI = urinary tract infections; HRT = hormone replacement therapy; PCT = Primary Care Trust.

\*The sample sizes (in the preceding three columns) can be multiplied by this value to obtain the samples required to derive quality scores at the PCT level.

†Not computed as sample too small.

sample needs to be eight times as large. This is because practices tended to meet either all or none of the pneumococcal indicators resulting in a between practice variance many times larger than that within practices, thereby increasing the cluster effect.

## DISCUSSION

### Summary of key findings

We found that conducting simultaneous quality assessment across multiple conditions is a highly resource intensive exercise, even when using indicators with a high level of face validity. A major practical barrier is the poor availability of data in general practice records. This would be partially overcome if practices made greater use of computerised records. However, in addition, there are significant technical barriers to conducting multi-condition assessments, particularly in calculating aggregate or summary scores for each condition or each provider of care (practice or PCT). There is no perfect way of aggregating scores to a higher level. The principal problem is determining the relative contribution of different indicators to the condition level score or the relative contribution of different conditions to a provider level score. The process is complicated by the fact that not all of the indicators are applicable to all patients with a given condition.

We have shown that it is very difficult to make reliable judgements about quality of care at the level of individual indicators. This is because very large samples are required, even if it is possible to sample medical records by condition, and this is not feasible at present for most conditions in many UK practices. Assessments are more reliable if indicators are aggregated to give condition level scores. Inspection of condition level scores in our sample of practices confirms the published evidence that there is significant variation in quality of care.<sup>19 20</sup> However, even at the condition level it may still not be feasible to conduct reliable assessments based on a random sample of records and it would be necessary to be able to identify the records of patients with the relevant diagnosis. Aggregation across a number of conditions to practice or PCT levels may be more feasible, although in the latter case cluster effects need to be taken into account.

### Critique of the approach adopted

One of the reasons for conducting this demonstration project was to explore the inherent barriers to conducting a comprehensive quality assessment of care. However, there were also conceptual and practical limitations to the approach that we adopted. In conceptual terms, our assessment of clinical care was limited in scope. Packaging the clinical components of primary care into simple technical processes risks marginalising the importance of the quality of interpersonal care.<sup>21</sup>

In practical terms, we were aware that our approach only assessed the care that had been recorded and it is possible that these data were not an accurate record of the care that was actually given. However, the method that we used to develop the set of indicators made the recording of relevant data an explicit element of the quality of care that was provided. We made no attempt to adjust the quality of care scores for factors outside the control of practices, such as population demographics, patient choice, or disease severity. Finally, the complexity of data collection using our approach dictates that inter-rater reliability checking is of great importance. We only had the resources to check this for five conditions. While the results for four of these were acceptable, the inter-rater reliability for one condition (depression) was poor and we are not sure why this was the case. This poor reliability was despite using research

### Key messages

- Conducting simultaneous multi-condition quality assessments in general practice is a highly resource intensive activity.
- There are significant problems in relation to poor data quality, low prevalence of some indicators/conditions, and methodological complexities in calculating summary scores for conditions and care providers.
- Random sampling and manual data extraction cannot be relied upon to assess quality of care using indicators. While some barriers may be overcome by the use of computerised records and sophisticated information management systems, these innovations alone will not solve more fundamental problems such as data quality, reliability, and the accuracy of records in reflecting the care actually provided to patients.
- Further research is required to develop methods for calculating summary scores and to compare indicator assessment based on computer codes and electronic datasets with that of random sampling.

nurses with considerable experience to extract data from the medical records.

### Implications for future research, policy, and practice

Individual quality indicators are useful tools for primary care teams to use within their practices. However, if valid and reliable external judgements are to be made about the relative quality of care of different practices, then individual indicators are much less useful and composite scores at condition or even modality levels (acute, chronic, preventive) may need to be used. The methods used to create composite scores are in their infancy and more work is required to move beyond the simple approach that we describe in this study to examine the merits and problems of different methods.

There are high expectations among policy makers and managers that data about quality of care will play a significant role in improving patient care and health system efficiency. This study shows that in general practice this ambition is well ahead of the science that is needed to support it—random sampling of records and manual data extraction cannot be relied upon to assess quality of care using indicators. The principal barrier is the poor quality of data and information management systems in general practice. As general practices become increasingly “paperless” and more accurate diagnostic and symptomatic computer codes are developed, research will be required to compare indicator assessment based on computer codes and electronic datasets with that of random sampling. Nevertheless, even the development of sophisticated information management systems is not likely to address the fundamental issue of data quality. How accurate are the data codes and how consistently are they used across general practices? And the question at the heart of all quality assessment exercises using either manual or computerised medical records—how accurately does the record reflect the care that has been provided? This intractable problem suggests that the use of clinical indicators should only form one part of a quality assessment process.

### ACKNOWLEDGEMENTS

The authors thank the Nuffield Trust for funding and supporting this project, Robert Brook, Paul Shekelle, Beth McGlynn and John Adams at RAND Health, California for their expert advice, and the staff in the sample practices for their support and practical help. The

National Primary Care Research and Development Centre is funded by the UK Department of Health. The views expressed are those of the authors and not necessarily of the funding bodies or the project advisors.

The project was devised by MM and all authors contributed to the design. SK and SKW collected the data and DR, SK and SC conducted the statistical analysis. SK wrote the first draft of the paper and all authors contributed to subsequent drafts. MM is the guarantor of the paper.

#### Authors' affiliations

**S A Kirk, S M Campbell, S Kennell-Webb, D Reeves, M O Roland, M N Marshall**, National Primary Care Research and Development Centre, University of Manchester, Manchester M13 9PL, UK

Conflicts of interest: none declared.

#### REFERENCES

- 1 **Donabedian A.** *Explorations in quality assessment and monitoring Volume 1: The definition of quality and approaches to its assessment.* Michigan: Health Administration Press, Ann Arbor, 1980.
- 2 **Brook RH, McGlynn EA, Cleary PD.** Measuring quality of care. *N Engl J Med* 1996;**335**:966–70.
- 3 **Campbell SM, Roland MO, Shekelle PG, et al.** Development of review criteria for assessing the quality of management of stable angina, adult asthma, and non-insulin dependent diabetes mellitus in general practice. *Qual Health Care* 1999;**8**:6–15.
- 4 **McColl A, Roderick P, Gabbay J, et al.** Performance indicators for primary care groups: an evidence based approach. *BMJ* 1998;**317**:1354–60.
- 5 **Marshall M, Campbell S, Hacker J, et al.** *Quality indicators for general practice: a practical guide to clinical quality indicators for primary care health professionals and managers.* London: Royal Society of Medicine Press, 2001.
- 6 **Roland M, Marshall MN.** General practice in an age of measurement. *Br J Gen Pract* 2001;**51**:611–2.
- 7 **McCormick A, Fleming D, Charlton J.** *Morbidity statistics from general practice. Fourth national study 1991–1992.* London: HMSO, 1995.
- 8 **Carroll K, Majeed A.** Co-morbidity associated with atrial fibrillation. *Br J Gen Pract* 2001;**51**:884–91.
- 9 **Starfield B.** New paradigms for quality in primary care. *Br J Gen Pract* 2001;**51**:303–9.
- 10 **Campbell S, Braspenning J, Hutchinson A, et al.** Research methods used in developing quality indicators in primary care. *Qual Saf Health Care* 2002;**11**:358–64.
- 11 **Willis J.** *The paradox of progress.* Oxford: Radcliffe Medical Press, 1995.
- 12 **Heath I.** *The mystery of general practice.* London: Nuffield Provincial Hospitals Trust, 1995.
- 13 **Smith P.** On the unintended consequences of publishing performance data in the public sector. *Int J Pub Admin*, 1995;**18**:277–310.
- 14 **Marshall M, Roland M.** The new contract: renaissance or requiem for general practice? *Br J Gen Pract* 2002;**52**:531–2.
- 15 **Brook RH, Chassin MR, Fink A, et al.** A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986;**2**:53–63.
- 16 **Department of Health.** *Shifting the balance of power: securing delivery.* London: Department of Health, 2002.
- 17 **McGlynn EA, Kerr EA, Asch SM.** New approach to assessing clinical quality of care for women: the QA tool system. *Women's Health Issues* 1999;**9**:184–92.
- 18 **Kerr EA, Asch S, Hamilton EG, et al.** *Quality of care for general medical conditions: a review of the literature and quality indicators.* Santa Monica, CA: RAND Health, 2000.
- 19 **Schuster MA, McGlynn EA, Brook RH.** How good is the quality of health care in the United States? *Milbank Q* 1998;**76**:517–63.
- 20 **Seddon ME, Marshall MN, Campbell SM, et al.** Systematic review of studies of quality of clinical care in general practice in the United Kingdom, Australia and New Zealand. *Qual Health Care* 2001;**10**:152–8.
- 21 **Campbell SM, Roland MO, Buetow S.** Defining quality of care. *Soc Sci Med* 2000;**51**:1611–25.

Want to know more?

Data supplements

Limited space in printed journals means that interesting data and other material are often edited out of articles; however, limitless cyberspace means that we can include this information online.

Look out for additional tables, references, illustrations.

[www.qshc.com](http://www.qshc.com)