

# Can health care quality indicators be transferred between countries?

M N Marshall, P G Shekelle, E A McGlynn, S Campbell, R H Brook, M O Roland

*Qual Saf Health Care* 2003;12:8–12

See end of article for authors' affiliations

Correspondence to: Professor M N Marshall, National Primary Care Research and Development Centre, University of Manchester, Williamson Building, Oxford Road, Manchester M13 9PL, UK; martin.marshall@man.ac.uk

Accepted for publication 8 July 2002

**Objective:** To evaluate the transferability of primary care quality indicators by comparing indicators for common clinical problems developed using the same method in the UK and the USA.

**Method:** Quality indicators developed in the USA for a range of common conditions using the RAND-UCLA appropriateness method were applied to 19 common primary care conditions in the UK. The US indicators for the selected conditions were used as a starting point, but the literature reviews were updated and panels of UK primary care practitioners were convened to develop quality indicators applicable to British general practice.

**Results:** Of 174 indicators covering 18 conditions in the US set for which a direct comparison could be made, 98 (56.3%) had indicators in the UK set which were exactly or nearly equivalent. Some of the differences may have related to differences in the process of developing the indicators, but many appeared to relate to differences in clinical practice or norms of professional behaviour in the two countries. There was a small but non-significant relationship between the strength of evidence for an indicator and the probability of it appearing in both sets of indicators.

**Conclusion:** There are considerable benefits in using work from other settings in developing measures of quality of care. However, indicators cannot simply be transferred directly between countries without an intermediate process to allow for variation in professional culture or clinical practice.

Quality indicators are increasingly used to facilitate regulation, ensure accountability, and improve quality.<sup>1–3</sup> In recent years there has been considerable interest in using high level indicators to compare the performance of different health systems.<sup>4–5</sup> However, developing lower level clinical quality indicators is an expensive and time consuming process, and there is currently little evidence to suggest that the process can be facilitated by transferring indicators developed for the health system in one country to another country. Researchers in the United States have led the way in the development and use of quality indicators for health care. In principle it might make sense for other countries to use this expertise rather than to develop their own indicators de novo, and there is some experience of doing this from the Maryland Hospital Indicator project.<sup>6</sup> However, it is unclear whether such a direct technology transfer is always appropriate or desirable.<sup>7–8</sup>

The aim of this study was to develop a set of quality indicators for UK general practice, aimed primarily at clinicians for quality assessment and improvement purposes. The indicators were based on work carried out at RAND in the US, one of the leading organisations in the field. We outline the methods used to develop a set of indicators in the US and describe a demonstration project that used the same methods to develop indicators for use in the UK. By comparing the outcomes of the two parallel processes, we draw conclusions on the extent to which quality indicators and their associated technologies can be transferred between countries.

## METHODS

### Selection of conditions

The RAND quality indicators (termed the "QA tools"<sup>9–12</sup>) address 58 clinical areas and were developed between 1995 and 1999 using a modification of the RAND/UCLA method of systematically combining evidence with expert opinion.<sup>13</sup> This process involves the development of a draft set of indicators following a comprehensive review of the literature and

then the rating of these indicators by experts for their validity as measures of quality. The process is described in detail in box 1.

In this study we wanted to develop quality indicators for the most common conditions seen in British general practice. We chose 19 of the 58 US conditions, concentrating on those conditions with the highest consultation rates in general practice based on data from the UK National Morbidity Survey<sup>14</sup> (table 1). The UK indicator set was developed between September 1999 and April 2001.

### Literature reviews

At the time of the UK study the US literature reviews from which the indicators (and the rating of the indicators) are

### Box 1 Steps in the RAND method<sup>13</sup> of developing quality indicators

- (1) Comprehensive literature reviews are commissioned for each of the conditions by experts in the field and a preliminary set of indicators are recommended by the author on the basis of the literature review and after consulting with clinical experts.
- (2) Expert clinicians are recruited from professional organisations and invited to join panels for a two stage process to rate the indicators.
- (3) Draft indicators and literature reviews are sent by post to the panel members who rate them in terms of their validity as measures of quality and the feasibility of collecting the data specified by the indicators. The panel members give each indicator two ratings on a continuous scale of 1–9.
- (4) First round scores are fed back to panellists for a second round of scoring in a two-day face to face panel meeting. Each panellist is told his or her own score and the mean score and distribution across the whole panel. All indicators are discussed, modified where necessary, and re-scored.
- (5) Second round scores are used to select only those indicators rated highly for validity and for feasibility by the panel members.

**Table 1** Numbers of US indicators which were in the final UK indicator set

	No of indicators in US set	No of identical or near identical indicators in UK set	Main reason for discrepancy
Acne	4	1	2 indicators relate to iso-tretinoin treatment which can only be prescribed by specialists in the UK
Allergic rhinitis	4	4	
Asthma (excluding indicators for hospitalised patients)	17	5	Likely difference in panel process (see text)
Cervical screening	7	3	More frequent monitoring in US
Contraceptive treatment	3	2	
Coronary artery disease (excluding indicators for hospitalised patients)	6	6	
Depression	17	9	Some differences in requirement for recording, also differences relating to available specialist resources
Diabetes	12	6	More frequent monitoring recommended by US panel
Dyspepsia and peptic ulcer disease (excluding procedures during/following endoscopy)	10	5	Likely difference in panel process (see text)
Headache	20	13	Different clinical practice: less biomedical approach in UK
Hormone replacement treatment	2	2	
Hypertension	12	8	More frequent screening recommended by US panel
Influenza immunisation <50 years of age	7	7	
Low back pain	14	8	
Osteoarthritis	7	4	
Otitis media in children	3	1	
Respiratory tract infection	11	3	Majority of differences relate to increased propensity of US physicians to use antibiotics or to investigate
Urinary infection	18	10	More emphasis on investigation and follow up in US. Longer courses of antibiotics in US
Total	174	98	

derived were 3–5 years old. We therefore decided to update the reviews, using leading UK primary care researchers with a specific interest in each condition. In general the UK reviews differed from those from the US through the inclusion of additional papers judged relevant to UK primary care rather than through the inclusion of more up to date evidence. The UK reviewers were asked to examine the US quality indicator set and, where appropriate on the basis of their experience and the modified literature reviews, to remove or suggest additional quality indicators.

### Selection of expert panels

Two panels of UK general practitioners were recruited by inviting participation from all 196 doctors who had been awarded the Fellowship by Assessment (FBA) of the Royal College of General Practitioners, the highest quality award for general practitioners in the UK; 75% agreed to take part. From these we purposefully selected nine members for each of two panels on the basis of sex, time since qualification, practice characteristics, and geographical location.

### Rating the indicators

Reproducing the RAND/UCLA process used to develop the US indicators, the panel members were sent the draft UK set of indicators and their supporting literature reviews by post. Panellists were asked to rate each indicator on a continuous 9-point scale in terms of its validity as a quality indicator and the “necessity to record” the information in a patient record (box 2).<sup>10</sup> One of the panels rated the indicators for ten of the conditions and the other for nine of the conditions.

The first round scores for validity and necessity to record were fed back to each of the panels at two-day face to face meetings chaired by PS/MM and MM/SC. All the indicators were discussed, first round scores for validity and neces-

sity to record were presented, the wording of the indicators was modified where necessary, and each indicator was then re-rated.

### Comparison of US and UK indicators

The second round ratings were used to select the set of UK indicators for comparison with the US parent indicators. To ensure comparability with the US set for the analysis presented in this paper, we used the same cut off points for US

#### Box 2 Definitions of validity and necessity to record used for rating indicators by expert panels<sup>11</sup>

An indicator is defined as valid if:

- there is adequate scientific evidence and professional consensus to support it;
- there are identifiable health benefits to patients who receive the care specified by the indicator;
- panel members consider that doctors with higher rates of adherence to the indicator would be considered as providing a higher quality of service;
- most factors determining adherence to the indicator are under the control of the doctor.

An indicator is considered as “necessary to record”\* if:

- in the opinion of the panel members, general practitioners in the UK should record this item of information in the medical record, such that absence of recording was a manifestation of poor quality care in itself.

\*This definition was different from that used by the US research team which used the term “feasibility” to describe the likelihood of finding data in an average medical record. In practice the terms address similar issues.

**Box 3 Examples where there was not an exact match between indicators but which were classified as “near equivalent” in the US and UK sets**

**Diabetes**

US: Type 2 diabetics who have failed dietary treatment should receive oral hypoglycaemic treatment.

UK: If the HbA1c level of a diabetic patient is measured as >8%, the following options should be offered 6 months apart: change in dietary or drug management, explanation of the raised test, or written record that a higher level is acceptable.

Note: Key common point is that records need to indicate action taken where glycaemic control is poor; the US indicator would require more detailed operationalisation before it could be applied to medical records.

**Nasal congestion**

US: If topical decongestants are prescribed, duration of treatment should be no longer than 4 days.

UK: If topical decongestants are prescribed, patients should be advised that duration of treatment should be no longer than 7 days.

Note: Key common point is prevention of rhinitis medicamentosa; the difference between 4 and 7 days is not of great clinical significance in this context.

and UK sets (validity scores  $\geq 7$ , necessity to record scores  $\geq 4$ ) without disagreement within the panel (three or more of the nine ratings for an indicator being in both the top and bottom third of scores<sup>10</sup>). The UK indicators that we finally published<sup>15</sup> were selected on the basis of more stringent cut offs for validity ( $\geq 8$ ) and necessity to record ( $\geq 6$ ). Although acute diarrhoea in children was included in both indicator sets, we excluded it from this comparison as the use of different age cut offs by the panels made valid comparison impossible. We also excluded US indicators that related to hospital based procedures such as during admission after

acute myocardial infarction as UK primary care physicians do not normally provide inpatient care. The comparisons reported in this paper are therefore based on 18 conditions. For each indicator in the US set we identified whether there was an exact or near equivalent indicator in the UK set. Examples of indicators which were classified as “near equivalents” are shown in box 3. Table 2 gives examples of indicators which were different.

**RESULTS**

Ninety eight of the 174 US indicators (56.3%) had near or exact equivalents in the final UK set (table 1). US indicators could have been discarded either by the UK reviewers if they were clearly not relevant to UK general practice, or during the panel process, or as a result of the panel scores; we did not attempt to distinguish between these as the purpose of this analysis was to compare the overall outcome of the two processes.

For the 159 indicators in the US set for which it was possible to classify strength of evidence, there was a slight but non-significant relationship between the strength of evidence for an indicator and the probability of the indicator having a near or exact equivalent in the final UK set: level 1 evidence, mainly randomised controlled trials, 64.3% (18/28); level 2, mixed evidence, 58.9% (10/17); level 3 evidence, mainly expert opinion, 54.4% (62/114),  $\chi^2$  0.96 df=2; test for linear trend not significant (p=0.34).

A few of the conditions are described in more detail below to illustrate the main reasons for the differences between the US and the UK indicators.

**Asthma**

Of the 17 indicators in the US set there were exact or near equivalents for five in the UK set. This discrepancy appeared in part to be related to the approach the panels had taken to the indicators rather than to fundamental differences in management. In particular, the US panel had eight indicators relating to care for acute exacerbations in the physician's office

**Table 2** Examples of differences between the indicators in the US and UK sets

Examples of differences between US and UK indicators	US indicator (strength of evidence as assessed by US panel)	UK indicator	Comment
Diabetes	Patients with a diagnosis of diabetes (type 1 or 2) should have glycosylated haemoglobin or fructosamine measured every 6 months (1/3)	HbA1c levels should be checked in diabetic patients at least every year	More frequent monitoring recommended by US panel
Depression	Patients who have suicidality should be admitted if they have psychosis, current drug or alcohol dependency (3)	No equivalent indicator	The second of the US indications is unrealistic in most NHS settings, because of lack of inpatient beds and drug and alcohol treatment facilities. Such patients are often managed in the community.
Headache	Recurrent moderate or severe tension headache should be treated with a trial of tricyclic antidepressant agents if there is no medical contraindication (1)	No equivalent indicator	Tricyclics not routinely used for this purpose in the UK. UK physicians are possibly less interventionist and may therefore draw different conclusions when there is only limited trial evidence
Hypertension	Systolic and diastolic blood pressure should be measured on patients presenting for care at least once each year (3)	All adults over 25 should have had their blood pressure measured in the previous 5 years	Much more frequent screening recommended by US panel
Osteoarthritis	Patients with incident symptoms of hip OA should be offered an A/P film of the affected hip (3)	Patients with a new diagnosis of osteoarthritis should not be x-rayed routinely	Greater reliance on radiological investigation by US physicians
Respiratory tract infection	Patients with sore throat and fever, tonsillar exudate and anterior cervical adenopathy should receive immediate treatment for presumed streptococcal infection (2)	For patients with sore throat who are systemically unwell and/or have 3 out of 4 Centor criteria, the doctor should discuss the pros and cons of treatment	More aggressive approach to antibiotic treatment by US physicians.

compared with only four in the UK set. Two US indicators related to theophylline which is rarely used for asthma in the UK.

### Cervical screening

Of the seven indicators in the US set there were exact or near equivalents for three in the UK set. The main reasons for the discrepancy were lower thresholds for action in the US set. Examples of these included shorter routine smear interval (3 years in US *v* 5 years in UK) and lower threshold for colposcopy (two moderately abnormal smears in the US *v* three in the UK). However, in one instance the UK panel recommended earlier action (repeat smear of colposcopy after moderately abnormal smear within 1 year in US *v* 6 months in UK).

### Coronary artery disease

Of the 18 indicators in the US set, 13 related to hospitalised patients and so were not appropriate to UK primary care. The remaining five had exact or near equivalents in the UK set.

### Depression

Of the 17 indicators in the US set there were exact or near equivalents for nine in the UK set. Two of the differences appeared to relate to expected standards of documentation. For example, the US panel specified enquiry about current medication when depression was diagnosed. The UK panel rated this as valid, but not necessary to record, probably because this information would already have been in the record of the UK primary care practitioner. In one case (see table 2) resources do not exist in the UK to provide the care recommended by the US indicator. UK panelists would therefore be unlikely to regard admission under these circumstances as necessary. Most of the other discrepancies related to a higher level of detail in the US set.

### Diabetes

Of the 12 indicators in the US set there were exact or near equivalents for six in the UK set. The most common reason for this discrepancy was more frequent monitoring specified by the US panel (table 2).

### Dyspepsia

Of the 10 indicators in the US set there were exact or near equivalents for five in the UK set. All but one of the discrepancies related to actions to be taken following endoscopy. The UK panel therefore appears to have concentrated more on initial diagnosis and treatment.

### Headache

Of the 20 indicators in the US set there were exact or near equivalents for 13 in the UK set. The UK set indicates a generally more conservative approach—for example, no indication for MRI based solely on the severity of headache, shorter list of drugs for first line prophylaxis of migraine, and no trial of tricyclic drugs required for moderate to severe tension headache (table 2). However, the UK indicators did require documentation of psychosocial history for patients with new headache (which the US indicators did not). This is consistent with the hypothesis that UK primary care physicians take a less biomedical (as opposed to biopsychosocial) approach to the problem.

### Hypertension

Of the 12 indicators in the US set there were exact or near equivalents for 10 in the UK set. For the other two, one related to more frequent screening required by the US panel (table 2) and, in the other (physical examinations to be carried out on new hypertensive patients), the UK panel gave lower validity scores than did the US panel.

### Osteoarthritis

Of the seven indicators in the US set there were exact or near equivalents for four in the UK set. There was a major difference in the interpretation of the benefits of radiological investigation (table 2)

### Otitis media

Of the three indicators in the US set there were exact or near equivalents for two in the UK set. There was no equivalent in the UK set for the US indicator that required at least 10 days of treatment with antibiotics for acute otitis media in children aged 1–3. Indeed, the UK set contained an indicator advising that, for children over 2 years, antibiotics should not be prescribed unless there were complications or persistent pain or discharge (over 72 hours).

### Respiratory tract infection

Of the 11 indicators in the US set there were exact or near equivalents for three in the UK set. There appear to be major differences in approach of the two countries to treatment and management of upper respiratory tract infection (table 2). In particular, UK physicians appear less likely to believe that there is an evidence base for aggressive treatment of streptococcal disease.

### Urinary tract infection

Of the 18 indicators in the US set there were exact or near equivalents for 10 in the UK set. In the US set there was generally more emphasis on pretreatment urine culture (required for a wider range of existing conditions) and on follow up investigation. The US panel also generally recommended longer courses of treatment (maximum 7 days for lower tract infection and minimum 10 days for upper tract infection *v* 5 and 7 days, respectively, in the UK). There was also a recommendation for a drug by the US panel (trimethoprim/sulfamethoxazole combination) which is regarded as obsolete in the UK (where trimethoprim alone is the recommended drug).

Further details are included in table 1 for allergic rhinitis, influenza immunisation, acne, low back pain, contraceptive treatment, and hormone replacement therapy. In these conditions there was generally close agreement between the UK and US set of indicators.

## DISCUSSION

There were considerable benefits in using US indicators as a starting point for developing a set of quality indicators for the UK, despite the need to replicate the US development process in order to produce contextually valid indicators for the UK. Collaboration between the UK and the US research teams resulted in new insights for researchers from both countries into the different purposes of quality indicators and into the impact of cultural and organisational factors on quality indicators. Fifty six percent of indicators in the US set had exact or near equivalents in the UK set. These indicators could be used as a basis for comparing quality of care in the two countries although, as noted above, in the final set of UK indicators more stringent cut offs for validity and necessity to record were chosen than in the comparative analysis reported in this paper.

Although we have focused on the presence or absence of US indicators in the UK indicator set as a means of assessing the applicability of the former in a second country, there were also indicators which appeared in the UK set alone. Sometimes these were clearly due to differences in the panel process—for example, detailed indicators on the management of hypertension in patients with angina in the ischaemic heart disease set—and sometimes they were related to the different health-care context—for example, requirement for registers of patients with angina diabetes and hypertension in the UK sets alone.

### Key messages

- Quality indicators are being developed independently in many different countries and are increasingly being used to draw comparisons between countries.
- The appropriateness and desirability of transferring indicators between countries and of using a common set of indicators for international comparisons has received little attention.
- This study suggests that there is considerable scope for countries to collaborate in the development of quality indicators, but there will always be important contextual differences between countries which mean that indicators cannot be transferred from one country to another without going through a process of modification.

We have focused on differences in professional practice in the results reported here. However, there are a number of additional reasons for differences between the two sets of indicators which we have not analysed in detail. Firstly, the literature reviews were different, with the UK reviews being more comprehensive and focused on primary care evidence than the US reviews—for example, the UK reviews were on average 14.7 pages long and contained 69 references while the US reviews were on average 8.3 pages long and contained 30 references. In addition, there may have been differences which related to the selection of indicators for scoring by the panels, the composition of the panels, and the conduct of the panel meetings. Finally, the reproducibility of the panel process is not perfect, although the reliability of panels rating the same set of indicators is generally regarded as acceptable.<sup>16 17</sup> In our analysis we have principally concentrated on differences which related to differences in clinical practice in the two countries.

This study has significant implications for other developed countries that plan to use indicators to improve quality and manage performance. We believe that there is considerable scope for countries to collaborate in the development of quality indicators, particularly countries with similar health systems such as the UK and The Netherlands. Nevertheless, there will always be important contextual differences between countries which mean that indicators cannot be transferred from one country to another without going through a process of modification.

### ACKNOWLEDGEMENTS

This project was commissioned and funded by the Nuffield Trust, London, UK. The National Primary Care Research and Development Centre is principally funded by the UK Department of Health.

### Authors' affiliations

**M N Marshall, S Campbell, M O Roland**, National Primary Care Research and Development Centre, University of Manchester, Manchester M13 9PL, UK  
**P G Shekelle, E A McGlynn, R H Brook**, RAND Health Program, Santa Monica, CA, USA

This study was devised by MM and all authors contributed to the design. MM, SC and MR managed the development of the UK indicators. MM and MR wrote the first draft of the paper and all authors contributed to subsequent drafts. MM and MR are the guarantors of the paper.

Conflicts of interest: none.

### REFERENCES

- 1 **Donabedian A**. *Explorations in quality assessment and monitoring 1. The definition of quality and approaches to its assessment*. Ann Arbor: Health Administration Press, 1980.
- 2 **Brook RH**, McGlynn EA, Cleary PD. Measuring quality of care. *N Engl J Med* 1996;**335**:966–70.
- 3 **Baker R**. Managing quality in primary health care: the need for valid information about performance. *Qual Health Care* 2000;**9**:83.
- 4 **World Health Organisation**. *The World Health Report 2000. Health systems: improving performance*. Geneva: World Health Organisation, 2000.
- 5 **Smith PE**. *Measuring up: improving health systems performance in OECD countries*. Paris: OECD, 2002.
- 6 **Thomson RG**, McElroy H, Kazandjian VA. Maryland hospital quality indicator project in the United Kingdom: an approach for promoting continuous quality improvement. *Qual Health Care* 1997;**6**:49–55.
- 7 **McKee M**, Figueras J. Comparing health systems: how do we know if we can learn from others? *J Health Serv Res Policy* 1997;**2**:122–5.
- 8 **Mulligan J**, Appleby J, Harrison A. Measuring the performance of health systems. *BMJ* 2000;**321**:191–2.
- 9 **McGlynn E**, Damberg C, Kerr E, et al. *Quality of care for children and adolescents*. Santa Monica, CA: RAND Health, 2000.
- 10 **McGlynn EA**, Kerr E, Damberg C, et al. *Quality of care for women: a review of selected clinical conditions and quality indicators*. Santa Monica, CA: RAND Health, 2000.
- 11 **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.
- 12 **Kerr EA**, Asch S, Hamilton EG, et al. *Quality of care for cardiopulmonary conditions: a review of the literature and quality indicators*. Santa Monica, CA: RAND Health, 2000.
- 13 **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.
- 14 **McCormick A**, Fleming D, Charlton J. *Morbidity statistics from general practice. Fourth national study 1991–1992*. London: HMSO, 1995.
- 15 **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.
- 16 **Shekelle PG**, Kahan JP, Bernstein SJ, et al. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med* 1998;**338**:1888–95.
- 17 **Naylor CD**. What is appropriate care? *N Engl J Med* 1998;**338**:1918–20.