MEASURING GENERAL PRACTICE

A demonstration project to develop and test a set of primary care clinical quality indicators

Professors of General Practice

Martin N. Marshall

Martin 0. Roland

Research Fellows

Stephen M. Campbell

Sue Kirk, David Reeves

National Primary Care Research and Development Centre, Williamson Building
University of Manchester, Oxford Road, Manchester M13 9PL, UK

Vice President, RAND; Director, RAND Health
Professor of Medicine and Health Services, UCLA

Robert Brook

Associate Director RAND Health

Elizabeth A. McGlynn

Senior Scientist

Paul G. Shekelle

RAND Health, Main Street, Santa Monica, CA, USA

ISBN 1-8330-3425-1

Published by The Nuffield Trust
59 New Cavendish Street
London WIG 7LP

Telephone: 020 7631 8450
Facsimile: 020 7631 8451

E-mail: mail@nuffieldtrust.org.uk
Website: www.nuffieldtrust.org.uk

Charity Number: 209201

Designed by Nicholas Moll Design
Telephone: 020 8879 4080

Printed by The Ludo Press Ltd
Telephone: 020 8879 1881

CONTENTS

ACKNOWLEDGEMENTS

This report has been commissioned as part of a collaboration between the Nuffield Trust, London and RAND, Santa Monica, California. The project on which the report is based was conducted in partnership with the National Primary Care Research and Development Centre (NPCRDC), based at the University of Manchester, UK. RAND is a non-profit institution that helps improve policy and decision-making through research and analysis. The NPCRDC is a multi-disciplinary centre for health services research in primary care, funded largely by the Department of Health.

We are indebted to the Trust's quality Steering Group for their guidance and to the Trustees for their financial support. In addition, we would like to thank the many individuals who contributed to the project, including Jenny Hacker who helped develop the indicator set, Sandra Kennell-Webb who contributed to the demonstration project, David Reeves and John Adams who provided statistical advice and Sarah Heyes who provided administrative support.

We hope that the report will be of value to primary care policy makers, managers and practitioners in the United Kingdom and have therefore used UK-English spelling.

The views expressed here are ours and not necessarily those of the Nuffield Trust or the Department of Health.

MM
SC
SK
DR
MR

EM
PS
RB

March 2003

FOREWORD

The development and use of objective measures of the processes and outcomes of health care is a fundamentally important part of improving our health system. Some medical specialities have recognised this for years, for example the analysis of morbidity data by epidemiologists, the use of physiological data by anaesthetists and the auditing of mortality rates by cardiac surgeons. For almost thirty years British general practice has been a leading light in the quality field, mostly through the activities of the Royal College of General Practitioners. However, much of this work has focused on formative educational programmes and GPs have been relatively slow to engage with the measurement agenda.

There are some good explanations for this apparent tardiness. General practice is more than a bio-medical discipline; it defines itself in terms of its holistic functions, such as continuity of care, co-ordination and patient advocacy. It is neither easy nor always desirable to reduce some of these complex activities into specific measurable items. In addition, much of the evidence on which quality indicators are based is derived from highly selected populations and may not be applicable to the kind of patients seen in general practice.

Despite this, it would be a mistake for general practice to ignore the high value that our society places on objectivity. General practice is a fundamentally important part of the National Health Service (NHS) in the UK - over 99 percent of the population are registered with a family doctor, 95 percent of consultations with the NHS are conducted in primary care and about 80 percent of these are dealt with by primary care teams with no involvement by hospital specialists. It is therefore essential that general practice plays a leading part in the modernisation agenda and engagement with objective measurement is one part of this process. Valid and reliable measures of quality serve to promote greater accountability of health professionals and organisations, act as a catalyst for quality improvement, as a vehicle for greater public and patient involvement and provide an opportunity for patients and purchasers to exercise choice. For these reasons, general practice is starting to rise to the challenge to develop measures which reflect the values of the doctors and nurses working in primary care.

This report describes a demonstration project which has made a significant contribution to this process. The project was initiated and supported by the Nuffield Trust and forms one part of a wider body of the Trust's work focusing on quality of care. In particular, the Trust

has supported ground-breaking work on the policy and practice of the public reporting of information about quality of care, launched with the publication of Dying to Know in 2000[1]. The public disclosure of information about primary care services is as important as that for the acute sector but the information that is to be released has to be valid and reliable. Using state-of-the-art techniques, this project has developed scientifically robust measures of the clinical processes of care for the most common conditions seen in general practice. To achieve this, the Trust brought together two of the world's leaders in the field, the RAND Health Group from Santa Monica, California in the USA and the National Primary Care Research and Development Centre, University of Manchester in the UK. This international collaboration has done more than simply transfer measurement technologies - the exchange of expertise and sharing of experiences has resulted in new insights into the policy and practice underlying the development and use of quality indicators in the primary care sector.

There are still many challenges facing us as we develop a range of quality indicators relevant to general practice. The indicators presented in this report and the description of the processes underlying their development make an important contribution to this work. The report is essential reading for primary care professionals, managers and policy makers as we rise to the enormous challenge of improving primary health care for patients.

John Wyn Owen CB
March 2003

EXECUTIVE SUMMARY

Aims

This report summarises a project that aimed to develop and test a comprehensive set of clinical quality indicators for use in British general practice. The study was commissioned by the Nuffield Trust with the aim of transferring expertise and specific measurement technologies from the USA to the UK. The project was conducted as a partnership between the National Primary Care Research and Development Centre, University of Manchester, UK and RAND Health, Santa Monica, California, USA.

Background

- The development of methods and tools to measure quality and performance in health care is regarded as a fundamental component of improving health services. The use of specific measures in primary care has tended to lag behind that in specialist services, though the demand for such 'quality indicators' from policy makers, managers, professionals and the public is growing.
- Quality indicators can be expensive and time-consuming to develop de novo. RAND Health is acknowledged as a leader in the field for almost three decades and in recent years has developed a set of clinical quality indicators for use in primary care settings. This project was devised to examine the utility of these US indicators in UK general practice.

Methods

The project was conducted in two phases: first, indicator transfer and adaptation, then indicator field-testing.

In the first phase, the US indicators for 19 of the most common conditions presenting in general practice, representing acute, chronic and preventative care, were examined by the UK team and by a UK primary care clinical expert for each condition. The literature reviews on which the indicators were based were updated and relevant community-based studies conducted in the UK were added. The indicators and the literature reviews

were then given to expert panels who rated them for their validity as measures of quality and the importance of recording information about them in the patient records. The two-stage rating process was based on the RAND/UCLA Appropriateness Method and resulted in a set of quality indicators with high face validity.

- In the second phase of the study, the highly rated indicators were field-tested on 1600 randomly selected patient records in 16 general practices belonging to two demographically contrasting English Primary Care Trusts. This required manual extraction of data from paper and electronic patient records.

Principal findings

- The expert panels rated 168 indicators for the 19 conditions as valid measures of quality. This represents the most comprehensive clinical quality assessment tool ever developed for UK general practice.
- Poor quality of data in general practice records, both in terms of its availability and accessibility, represents a significant obstacle to quality assessment in primary care.
- There was considerable variability in the condition prevalence and the indicator prevalence in the random set of medical records that were examined. As a result, relatively few data were collected for the less prevalent conditions or indicators relating to less common aspects of care.
- Higher quality scores were achieved for preventative care (mean score 61.95% of indicators achieved for 6 conditions) than for acute or chronic care (mean scores 54.5% of indicators achieved for 8 conditions and 60.0% of indicators achieved for 9 conditions respectively).
- The method by which quality scores are weighted may have an effect on the resulting summary scores. The weighting method should reflect the purpose for which the scores are being used.

Policy implications

- Transferring quality assessment technologies and specific tools between countries achieves considerable benefits. However, primary care quality indicators cannot be directly transferred between countries with different health systems, clinical practices and cultures and so an intermediate process of adaptation is required.
- There are good arguments for aggregating individual indicator scores to produce summary quality scores at the level of modalities of care, conditions, clinicians, practices or larger primary care organisations. The merits and problems of different methods of computing summary quality scores and the implications of weighting scores for importance or prevalence are examined in this report.
- Quality indicators, such as those developed in this project, have an important role to play in professionally-led improvement, performance management and user

involvement. However, such a disease-focused approach to measuring quality in general practice will inevitably miss out many important aspects of care, such as issues relating to access and inter-personal care. In addition, the tension between professional led quality improvement and managerially led quality assessment needs to be recognised and addressed.

- Without a major investment in information systems in the NHS, quality assessment will always be expensive, time-consuming and of questionable reliability and validity. Only good information technology will enable the improvement agenda to move forward.
- The clinical indicators developed in this project represents only one part of what general practitioners and primary care nurses do. A comprehensive assessment of quality would need to examine the quality of inter-personal care, the relationship between primary care and other services, patient experience of care, the organisation and delivery of services and their cost effectiveness. Many of these dimensions of care have received minimal attention to date. Future work in this field should focus on further developing specific measures in these domains and on understanding the relative importance of the domains and how they relate to each other.
- Further work is required to test and understand the scientific properties of these indicators in order to maximise their use in quality improvement and their application in performance management.

CHAPTER 1

The policy context

A doctor working in general practice in the National Health Service (NHS) thirty or forty years ago would hardly recognise the demands and expectations made on today's general practitioners (GPs). Within the memory of some practising doctors, most GPs worked in isolation, detached from peers, loosely accountable to administrators for reimbursement purposes only and implicitly trusted by their patients to be competent and to remain up to date. Judgements were rarely made about the quality of care provided by GPs. This was probably just as well, since there were few agreed criteria by which to judge whether the care provided was good or fell below acceptable standards.

Today's GPs operate in a very different environment - more accountable for their actions, more likely to work as one member of a Primary Health Care Team, more often judged by peers, by managers and by patients and less likely to be trusted simple because they are professionals.[1] There are many inter-related reasons for these dramatic changes, several of which are relevant beyond the boundaries of general practice. In part they are a consequence of the widespread de-professionalisation of society and the high expectations resulting from consumerism. There are now greater demands for information and this feeds, and is fed by, the increasing availability of computerised data and advances in methods of measuring quality of care.[2] Politicians' demands for efficiency increase the pressure on managers to judge whether doctors are providing value for money. Alongside these changes have been some high profile examples of poor quality practice that have dented the public's confidence in implicit professional self-regulation.[3]

These forces for change have led to calls from the public, politicians and professional bodies for evidence that acceptable standards of care are being delivered by those who work in the NHS. The United Kingdom (UK) government is driving change aligned to a strategic plan described in a series of White Papers[4 5] and outlined in the recent NHS Plan.[6] This new strategic focus has sometimes failed to acknowledge the considerable work that has been done on quality improvement in general practice over the last half a century, much of it led by the Royal College of General Practitioners.[7] However, these activities have tended to be fragmented, poorly coordinated and poorly sustained. Perhaps most importantly,

improvement in general practice has tended to be insufficiently guided by objective measures of quality. It is this measurement issue, using specific measures or quality indicators, that this project was designed to address.

The use of indicators in primary care is not a new phenomenon. In the 1980s performance indicators were mostly used as a means of driving economy and efficiency savings. In the 1990s high level public health indicators were used to examine variation and inequalities in health and health provision. Since 1997 quality indicators have been used increasingly to improve quality, as part of the clinical governance agenda which seeks to combine effectiveness and efficiency.[8] This interest has been reflected by the wide ranging debate about the advantages and problems of applying indicators[9] [10] and the publication of specific indicators using rigorous developmental processes.[11] [12] [13]

This demonstration project aimed to drive the measurement agenda forward by bringing together two world leaders in the field, the RAND Health Group from Santa Monica, California in the USA and the National Primary Care Research and Development Centre, University of Manchester in the UK. Both organisations have developed clear conceptual frameworks to define quality in primary care[14]-[15]-[16] and have developed a range of indicators for assessing the quality of clinical care provided in the community[12]-[17]-[18]-[19]-[20] Whilst individuals from the two organisations had worked together in the past, support from the Nuffield Trust allowed RAND Health and the NPCRDC to create a large-scale collaboration and to share their expertise. The result of this collaboration, the Global Assessment Project (GAP) is described in this report.

The following chapter provides an overview of the role of quality indicators in general practice.

CHAPTER 2

Quality indicators for primary care: an overview

This chapter defines quality indicators, examines their role in assessing and improving primary care in the UK, outlines the methods available for developing indicators and discusses the factors required to maximise their effectiveness in fostering quality improvement. It finishes by outlining some of the advantages and disadvantages of using quality indicators in a general practice setting.

Defining quality indicators

An indicator is a measurable item of care, which focuses upon some aspect of structure, process (inter-personal or clinical) or outcome.[14 21] Indicators need operationalising in order to be useful and are therefore often used to generate review criteria and standards. The relationship between guidelines, indicators, review criteria and standards is described in boxes 1 and 2.

There are various types of indicators. Activity indicators measure the frequency with which an event occurred (e.g. influenza immunisations). The term 'performance' indicator is sometimes used synonymously with quality indicator but it is possible to make inferences about performance without making inferences about quality (for example when referring to cost issues in isolation), so they should be differentiated from each other.[22] Quality indicators infer a judgement about the quality of care being provided.[23] Importantly, indicators indicate potential problems that might need addressing, usually manifested by low indicator scores, statistical outliers, or unexplained variation in care.

Box 1: Definitions of guideline, indicator, review criterion and standard[24]

Guideline:	Systematically developed statements to assist practitioner and patient decisions prospectively for specific clinical circumstances; in essence the "right thing to do".
Indicator:	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 in the quality, of care provided.
Review criterion:	Systematically developed statement relating to a single act of medical care that is so clearly defined it is possible to say whether the element of care occurred or not retrospectively in order to assess the appropriateness of specific health care decisions, services and outcomes.
Standard:	The level of compliance with a criterion or indicator. A target standard is set prospectively and stipulates a level of care that providers must strive to meet. An achieved standard is measured retrospectively and details whether a care provider met a pre-determined standard.

Box 2: Examples of a guideline, indicator, review criterion and standard[24]

Guideline recommendation:	If a blood pressure reading is raised on one occasion, the patient should be followed-up on two further occasions within X time.
Indicator:	Patients with a blood pressure of greater than 160/90 mm Hg should have their blood pressure re-measured within three months.
Indicator numerator:	Patients with a blood pressure greater than 160/90 mm who have had their blood pressure re-measured within three months
Indicator denominator:	Patients with a blood pressure greater than 160/90 mm
Review criterion:	If an individual patient's blood pressure was >160/90 was it re-measured within three months?
Target standard:	9<Jpercent of the patients in a practice with a blood pressure of greater than 160/90 mm Hg should have their blood pressure re-measured within three months.
Achieved standard:	80 percent of the patients in a practice who had a blood pressure of greater than 160/90 mm Hg had their blood pressure re-measured within three months.

Developing and applying quality indicators

When developing indicators, three preliminary issues need to be considered.[24] These relate to what is going to be measured, whose views are being represented and how the indicator can be developed.

What should be measured?

In order to answer this question it is first necessary to be clear about the working definition of quality. Having an explicit definition clarifies which aspects of care are, and equally importantly are not, being assessed using the indicators.

Quality is a difficult concept to define when applied to heath care and efforts to do so in a single sentence are often unhelpful. Most commentators have therefore focused on defining

quality in terms of the different dimensions of care rather than attempting to produce a generic definition. Perhaps the most widely known set of dimensions in the UK is that proposed by Maxwell, who described quality in terms of accessibility, effectiveness, efficiency, acceptability and equity.[25] Campbell *et al* suggested that quality could be seen simply in terms of access (can patients get to the care they need) and effectiveness (is it any good when they get there?).[16] Donabedian[21] suggested that quality can be seen in terms of structures (e.g. personnel, equipment, finances), processes (e.g. consulting, referral, prescribing) and outcomes (e.g. health status, user assessments).

There is, of course, a tension between these different dimensions of quality. Inevitably a trade-off has to be made - it is rarely possible to deliver good quality on all dimensions at the same time. General practitioners might reasonably argue that they could provide better clinical care if resources were put into, say, employing a practice nurse to run a diabetic clinic in their practice. However, this investment might better be made in providing extra community services for the mentally ill, or reducing waiting lists for an orthopaedic outpatient clinic. In addition there is a very real tension in general practice between the needs (and demands) of individual patients and those of the practice or local community as a whole.

This tension ensures that the debate about quality is often controversial. Whilst we recognise the danger of over-simplifying a complex issue, it is nevertheless helpful to start addressing quality in areas where there is clear evidence of problems[26] and where significant improvements are most likely to be achieved. For this reason, the GAP project focused specifically on the technical processes performed by general practitioners and practice nurses in delivering clinical care for the most common problems presenting to general practice. This pragmatic focus on clinical effectiveness is not meant to divert attention from other important dimensions of quality, particularly interpersonal care.

Whose views are being represented?

Different stakeholders have different perspectives about quality of care[27] and no one perspective is any more valid than another. These different stakeholders often disagree about what is important and this is why discussions about quality can be controversial. For example, managers tend to be more focused on efficiency, and increasingly on outcomes, whereas patients often relate quality to an understanding attitude and communication skills. It is important to be clear which stakeholder views are being represented when developing indicators. In the GAP study we have focused explicitly on clinical processes of care from a health professional perspective.

How can indicators be developed?

Quality indicators have been developed in a variety of different ways, some systematic and some non-systematic. The first and most common way, much used in the UK in the past, has been for a group of people to sit down together around a table and come up with suggestions. These are usually based on readily available information, such as referral rates, rather than on fundamental concepts or theories. This approach has the advantages of speed and simplicity but the disadvantage that resulting indicators may have little meaning to those who want to use them to improve the quality of clinical care.

A second approach is to base indicators on clinical guidelines, which are usually derived from a complex and sometimes poorly defined mixture of scientific evidence and professional opinion. This approach has been used extensively in The Netherlands and by some experts in the UK.[2829] A third approach is to focus purely on published evidence from randomised controlled trials. This "evidence-based" approach [11] has the advantage of producing rigorous and scientifically acceptable indicators but has two main disadvantages. First, it focuses on a very limited part of general practice - since so much of what is regarded as good quality care in general practice does not have (and probably never will have) experimental evidence to support it. Second, some people have questioned the applicability to individual patients of evidence derived from scientific trials on selected populations.

In response to this, in the GAP project we adopted a third approach, developed over 25 years ago by RAND and UCLA in California.[30] This approach recognises the importance of scientific evidence but is concerned with the application of this evidence to real clinical practice and with the significant gaps in the evidence applicable to some areas of practice. We therefore chose this rigorous method of combining scientific evidence with expert opinion in an attempt to produce a more comprehensive and useful set of quality indicators for British general practice. The method has been used extensively in both the US and the UK, and in both primary and secondary care.[31,32,12] Despite some criticisms,[33,34] it is generally regarded as the most rigorous and systematic way of combining expert opinion and scientific evidence.[35]

The characteristics of the RAND appropriateness method are summarised in box 3.

Box 3: Some of the key characteristics of the RAND appropriateness method

- **timeliness:** produces indicators within a relatively short time period.
- **systematic:** based on a systematic and comprehensive synthesis of available evidence.
- **knowledge based:** builds on the scientific literature by incorporating expert opinion.
- **explicit voting:** relies on equal input from each expert rather than requiring absolute consensus

The benefits and risks of using quality indicators in general practice

The benefits and problems of using quality indicators to assess and improve quality in general practice are summarised in boxes 4 and 5. The debate amongst health professionals has tended to concentrate on the problems and the negative consequences of using quality indicators. This is hardly surprising given the nature of the indicators in common use and the abuse of comparative data by some parties. However, more sophisticated methods are now being used to develop quality indicators and greater attention is being given to the scientific properties of the resultant indicators. This should result in the future in a more balanced debate about the risks and benefits.

Box 4: The benefits of using quality indicators[36]

Quality indicators can:

- Allow comparisons to be made between practices, over time or against gold standards. These comparisons can stimulate and motivate change.
- Facilitate an objective evaluation of a quality improvement initiative.
- Be used to ensure accountability and identify unacceptable performance.
- Stimulate informed debate about quality of care and level of resources.
- Focus attention on the quality of information in general practice.
- Help target resources to areas of greatest need.
- Be quicker and cheaper tools for quality assessment than other tools, e.g. peer-review.
- Inform purchasing decisions and planning of service agreements.

Box 5: The problems of using quality indicators[35]

Quality indicators may:

- Encourage a fragmented approach to an holistic and integrated discipline.
- Assess only easily measurable aspects of care and fail to encompass the more subjective aspects of general practice.
- Be based on dubious quality data and information that is difficult to access.
- Be difficult to interpret - for example apparent differences in care may relate more to random variation, case-mix or case severity, rather than real differences in the quality of care.
- Be expensive and time consuming to produce; the cost-benefit ratio of measuring quality of care is largely unknown.
- Encourage a blame culture and discourage internal professional motivation.
- Lead organisations to focus on measured aspects of care to the detriment of other areas and to concentrate on the short term rather than adopting a long-term strategic approach.
- Erode public trust and professional morale if deficiencies in the quality of care are highlighted.
- Encourage massaging or manipulation of the data by health professionals or organisations if the results of indicators are published.

Conclusion

Whilst there are risks to using quality indicators which should not be underestimated, they provide a useful means of both assessing and improving quality of care. Indicators need to be developed based on the best available evidence and be tested, as far as possible, for validity, reliability, acceptability, feasibility and sensitivity to change in their development and application. This will help to maximise the effectiveness of quality indicators in quality improvement strategies. Whilst this is most likely to be achieved when they are derived from rigorous scientific evidence, this is often unavailable for many aspects of care. Consensus techniques, which systematically combine evidence and opinion, therefore facilitate quality improvement because they allow a significantly broader range of aspects of care to be assessed and improved than would be the case if quality indicators were restricted to scientific evidence. The method that we chose to use in this project is a validated and respected consensus technique, the RAND/UCLA Appropriateness Method.

CHAPTER 3

Aims and objectives of the project

The aim of the project was to develop a set of clinical quality indicators for use in UK general practice, using the expertise and experience developed in this area by RAND Health. Our hope was that the partnership between the UK and US teams would maximise the chances of developing a high quality set of indicators, more quickly than would have happened if the UK team had worked in isolation.

The objectives were:

1. To review research on quality indicators for primary care developed in USA and consider the applicability of this work to UK general practice
2. To identify the most common and significant conditions presenting to UK general practice, focusing on those conditions for which researchers in the US have already conducted literature reviews and identified quality assessment indicators. To update and modify the US reviews for a UK context
3. To use the modified reviews to develop a list of draft indicators for use in the UK and then to conduct expert panels to refine and select the final set of indicators.
4. To conduct a demonstration project by using the final set of agreed indicators to audit care in practices in two Primary Care Groups/Trusts.
5. To disseminate the results and identify further research and development needs.

CHAPTER 4

Methods

1. Indicator development

Methods

The following section summaries the methods used to develop the indicator set, which have been described in detail elsewhere.[36]

Researchers from RAND Health have used the RAND/UCLA Appropriateness Method to develop primary care quality indicators for over 70 conditions presenting to US primary care physicians.[1718-1937] On reviewing these indicators it was clear that the structural and cultural differences between the US and UK health systems would result in significant problems if these indicators were applied directly to British general practice. We therefore used a modification of the RAND/UCLA Appropriateness Method to adapt the US indicators and develop new indicators that could be used for quality assessment in the UK.

The process was undertaken in six stages (box 6):

Box 6: Summary of the stages in the development of the indicators:

1. **Selection of conditions:** 19 common conditions were chosen for which indicators would be developed.
2. **Developing the literature reviews and preliminary sets of indicators:** literature reviews were commissioned for each of the conditions and preliminary indicators were developed.
3. **Selection of expert panels:** Experts in general practice were invited to join panels for a two stage process to rate the indicators.
4. **First round postal survey:** Draft indicators and literature reviews were sent to the panel members, who were asked to rate them in terms of their validity and the importance of recording the data.
5. **Second round panel meetings:** The first round scores were analysed and the results given back to panellists for a second round of scoring in a 2-day face-to-face panel meeting.
6. **Second round data analysis and drafting of final indicator set:** The second round scores were used to select only those indicators rated highly for validity and for necessity to record the information on which the indicator was based.

Stage 1: Selection of conditions

We chose to develop quality indicators for the most common clinical conditions presenting

to UK general practice, based on the most recent National Morbidity Survey in General Practice.[38] The 19 selected conditions, all of which had been reviewed as part of the RAND project, provide examples of acute, chronic and preventative care. Using national general practice morbidity data, we estimate that they represent about 60 percent of consultations in UK general practice.

Stage 2: Developing the literature reviews and preliminary sets of indicators

The literature reviews are an important part of the indicator development process because they encourage the experts to relate their opinions and experience to the available scientific evidence. New evidence-based reviews[36] were therefore commissioned from leading primary care researchers in the United Kingdom. They are not formal systematic reviews but represent comprehensive summaries of the available national and international literature, focusing specifically on evidence directly relevant to general practice in the UK.

At the same time as reviewing the literature, the reviewers were asked to propose a preliminary set of quality indicators for their condition. This set was based on the evidence, national guidelines and professional statements, and was influenced by the indicators developed by the RAND Health team. The indicator set and supporting data were presented to the expert panels in a structured format. An example is presented in table 1.

Table 1: Example of suggested indicator

	Indicator	Quality of evidence*	References	Benefits/summary
12	Patients with Coronary Artery Disease should be advised to take aspirin at a dose of 75-150mg/day (continued indefinitely) unless they have a contraindication	I	Yusuf era/., 1988; ATC, 1994; Khunti era/., 1999	Absolute reduction in vascular events of about 5%

*Quality of evidence

I based on evidence from randomised controlled trials

II-1 based on evidence from non-randomised controlled trials

II-2 based on evidence from cohort or case studies

II-3 based on evidence from multiple time series

III based on opinion or descriptive studies

Stage 3: Selection of expert panels

Panels of experts were then convened to rate the preliminary indicators. The definition of an "expert" in general practice is difficult. We wanted to ensure that the expert panel members were familiar with critical appraisal of scientific evidence but were grounded in the reality of daily general practice. We therefore decided to select panel members from the database of Fellows by Assessment of the Royal College of General Practitioners ("FBAs"), since this award, based on rigorous self and peer practice-based assessment, is generally regarded as the highest explicit standard attained by service general practitioners. All 196 FBAs in the UK in 1999 were invited to participate. Eighty-two

percent responded and 91 percent of these agreed to take part. Panel members were selected to represent the genders, different types of practice and geographical location and levels of clinical experience.

Two panels, each with 9 members, were formed. Each panel was allocated approximately half of the conditions to assess, as summarised in box 7.

Box 7: Selected conditions by panel

Panel A	**Panel B**
Coronary artery disease	Asthma
Hypertension	Depression
Diabetes mellitus	Osteoarthritis
Allergic rhinitis	Upper respiratory tract infections
Headache	Acute otitis media
Urinary tract infection	Diarrhoeal disease in children
Dyspepsia	Acne
Cervical screening	Low back pain
Immunisation	Family planning
	Hormone replacement therapy

Stage 4: First round postal survey

Panel members were sent, by post, the literature reviews and preliminary indicator sets for the conditions being rated by their panel. They were asked to rate all the indicators for their validity as indicators of quality and whether the information was important to be included in the patient's record ("necessity to record"). Each indicator was rated on a nine point continuous scale for validity and necessity to record, where 1 represented the lowest and 9 the highest rating.

The panellists were advised that an indicator should be considered valid if the following criteria were met:

- there was adequate scientific evidence and professional consensus to support it
- there were identifiable health benefits to patients who received the care specified by the indicator
- panel members considered that doctors or nurses with higher rates of adherence to the indicator would be judged as providing a higher quality service
- « most factors determining adherence to the indicator were within the control of the doctor or nurse

The panellists were told that ratings of 1 to 3 would mean that the indicator was not a valid measure of quality; 4 to 6 would mean that the indicator was of uncertain or equivocal validity and 7 to 9 would mean that it was considered to be a valid measure of quality.

An indicator would be considered as necessary to record if the following criteria were met:

- failure to document the information could be judged itself to be a marker for poor quality

- estimates of adherence to the indicator based on medical record data are likely to be reliable and unbiased.

The panellists were told that a rating of 1 to 3 would mean that the data should not have to be recorded in the medical record; 4 to 6 would mean that there is legitimate uncertainty about the need to record the data and 7 to 9 would mean that the information should be recorded in the patient's notes.

The panellists were also invited to suggest changes to the wording of the indicator if they thought appropriate. An example of a panellist's rating for one indicator is shown in table 2.

Table 2: Example of a first round postal rating

	indicator	validity	necessity to record
3.6	All diabetic patients should have an annual fundal examination	1 2 3 4 5 6 7 8 (9)	**1 2 3** 4(5)6 **7 8 9**

Stage 5: Second round panel meetings

Completed first round questionnaires were returned by all panellists to the research team. For the second round the ratings were summarised and fed back to the panel members in a face-to-face meeting. The purpose of the meeting was to discuss the first round ratings, prior to repeating the rating process. No indicators were dropped between rounds, irrespective of how they were rated in round one, in order to allow panellists the opportunity to discuss each indicator at the panel meeting. The panel meetings lasted for 2 days and were chaired by members of the research team. The chair's role was to facilitate discussion, focusing on the indicators for which there was wide variation in the ratings of different panellists. The panel members were not forced to reach consensus and were encouraged to rate as they saw fit after the discussion for each indicator.

The data were presented to the panellists as an anonymised overall distribution of the first round scores for all members, together with the first round score of the individual panel member. An example of the feedback and modification (shown in italics) to the indicator is shown in table 3. This shows the rating scale (1-9), the overall distribution of all panel members' scores in italics (i.e. six members gave a rating of "3" for validity) and this individual panel member's personal rating in bold (i.e. a validity score of "8" in this example).

Table 3: Example of feed back to panel members for second round ratings

	indicator	validity	necessity to record
10.8	Short-acting B2 agonist should be prescribed for symptomatic relief on an 'as required' basis *unless contraindicated or intolerant*	*7 7 6 7* 1 2 3 4 5 6 7 8 9	*3 6* 1 2 **3** 4 5 6 7 8 9

Stage 6: Second round data analysis and drafting of final indicator set

Only second round ratings were used to select the final set of indicators. The decision to select or reject indicators was based on their median validity and "necessity to record" scores and the level of agreement between the scores of panel members.

To be considered for inclusion in the final set, an indicator needed an overall median rating of greater than 7 for validity and greater than 6 for necessity to record, without disagreement within the panel. The RAND Health research team used the standard approach developed in the Appropriateness Method of greater than 6 for validity and greater than 4 for recording. Disagreement was defined in statistical terms as being when three or more of the nine ratings for any one indicator were in the one-to-three region and three or more in the seven-to-nine region.

The number and proportion of indicators rated by the panels as good measures of quality, by condition, is shown in table 4. A higher proportion of indicators for chronic and preventive care were rated valid than for acute care. This highlights the difficulty with making explicit judgements about quality for some aspects of general practice, particularly those that have a weaker evidence base.

Table 4: Proportion of indicators rated valid by condition

Condition	Number of indicators rated	Number rated valid	% rated valid
Asthma	**29**	25	86.2
Family planning and contraception	**7**	**6**	85.7
Cervical screening	**11**	**8**	72.7
Coronary heart disease	**18**	**13**	72.2
Dyspepsia and peptic ulcer disease	**13**	**9**	69.2
Acne	**9**	**6**	66.7
Hypertension	**43**	**28**	65.1
Immunisations	**31**	20	64.5
Acute otitis media	**5**	**3**	60.0
Urinary tract infection	**32**	**19**	59.4
Depression	46	25	54.4
Diabetes mellitus	**29**	**15**	51.7
Hormone replacement therapy	**18**	**9**	50.0
Acute diarrhoeal disease in children	25	**11**	44.0
Allergic rhinitis	**10**	**4**	40.0
Headache	**37**	**11**	29.7
Upper respiratory tract infection	27	**8**	29.6
Osteoarthritis	**23**	**6**	26.1
Acute low back pain	**22**	**3**	13.6

Further development of the indicator set

A total of 229 indicators for the 19 conditions were rated as valid measures of quality by the expert panels. However, this was just the first step in the process of indicator development. At the end of the panel process the indicators are conceptually valid measures of quality based on systematically combining scientific evidence and expert opinion. The next stage is to test the indicators in clinical settings in order to investigate their scientific properties. This phase of the study is described in the following section.

2. The demonstration project

The aim of this phase of the project was to examine the feasibility of using the indicator set in UK general practice. This was conducted in the following way:

(i) Developing and piloting the data extraction forms

The demonstration project commenced with the operationalisation of the indicators through the development of data abstraction forms (DAFs) and a detailed manual for each of the study conditions. The process was straightforward for some of the indicators, requiring just one or two questions, but was complex for others, requiring multiple questions to determine the appropriateness of the indicators to an individual patient.

Pilot work to test the DAFs and manuals was undertaken in four practices. The aim of this work was to test and further develop the DAFs and to provide training for the researchers in extracting data from medical records and in the use of practice-based computer systems. This stage of the demonstration project also provided an opportunity to refine the indicator set, with some indicators being removed due to the difficulties experienced in extracting the relevant data from the patient records. The inter-rater reliability of the extraction process was also piloted at this stage. This resulted in alterations to the data extraction forms to improve the clarity of question wording and ordering, the establishment of more explicit inclusion criteria for some conditions and the removal of some 'uncollectable' indicators. A total of 4 indicators were removed during these preliminary stages (table 5).

Table 5: Indicators removed during the preliminary stages of the project

Indicator Removed	Reason
Depression Children and adolescents (under 16 years of age) with major depression should be referred for a specialist assessment.	Insufficient numbers of children with 'major depression'. Difficulties in assessing the presence of 'major depression' from records.
URTI Rhinitis An antibiotic prescription should not be offered to patients' with uncomplicated infective rhinitis with symptoms of fewer than 14 days duration.	Rhinitis' found to be unused in medical records as a diagnostic label.
Acne For patients presenting with acne, enquiry should be made about previous treatment	Insufficiently specific
Family Planning Women requesting family planning advice should be offered the choice of all appropriate forms of contraception, which could be delivered either within the practice or by referral.	Insufficiently specific

(ii) Sampling Primary Care Trusts and practices

Two Primary Care Trusts, in the North West and the South West of England, were purposively selected on the basis of their different social deprivation profiles. A stratified random sample of ten practices from each Primary Care Trust was invited to participate in the study. Sixteen of these 20 practices (80%) agreed to take part (nine in one Primary Care Trust and seven in the other). Each participating practice received a £100 honorarium.

(iii) Data collection

The researchers contacted the practice managers at the participating practices to further discuss the study, in particular to clarify what it would involve for the practice, and to arrange a convenient time for data collection. Data collection occurred on a practice-by-practice basis over a period of approximately five months (December 2000 to April 2001).

At each practice a random sample of 100 patient records (both manual and computerised) were evaluated to examine whether or not the care determined by the indicators was provided. It was not possible to purposefully select patient records with each condition under study because of the generally poor quality of coding of general practice records in the UK - at the time of data collection even chronic disease registers were not considered to be sufficiently reliable.

Four different methods had to be used in the sample practices to generate a random list of 100 patients, reflecting differences in the availability of data. The quickest and easiest approach for those practices that were able to, was to generate such a list from their computer systems (three practices in the sample were able to produce such a list). The other three methods used a random number table to either: (1) generate computerised patient identification numbers (used in six practices); (2) to use in conjunction with a complete list of patients at a practice (used in seven practices); or (3) in the case of a non-computerised practice, to select which records to audit from the practice's record filing drawers.

Each set of records (manual and computerised for all patients) was examined to identify whether or not the patient had had a consultation within the previous 5 years and whether any of the study conditions were recorded in the notes. Records where patients had not been seen in this time period were discarded and another random set of records obtained. The study conditions had to be explicitly recorded in the records as the researchers were not expected to attempt to make a diagnosis on the basis of signs or symptoms recorded. In those records where none of the study conditions were recorded only the persons age and gender were noted.

Each applicable condition for each patient was then audited using the DAFs. The length of time taken to audit a set of records varied from a few minutes to over 1 hour. Some records had to be read and reread a number of times to obtain the data required. On average, approximately 15-20 records could be audited per day. Examining all 100 records in each practice took between seven and ten days, depending on the quality of the records. In total 1600 sets of records were reviewed over the five-month period.

In addition to the clinical data, background data were also collected about the practice (for example, list size, staff numbers, level of computerisation).

(iv) Data analysis

The first aim of the analysis was to test the feasibility of producing quality summary scores for individual conditions, modalities (acute, chronic or preventative care), practices and PCTs. The second aim was to investigate factors which might predict performance.

(a) The computation of summary scores

Data from the abstraction forms were entered into SPSS and then computed to produce a dataset consisting of a set of indicators and two pieces of information about each indicator in relation to each patient: (i) whether it applied to the patient or not; and (ii) where it did apply, whether it was met (i.e. the patient received the care indicated). Quality scores for individual indicators and conditions were produced across all patients and all practices. The condition scores were derived by dividing the total number of occasions that the indicators for that condition were met by the total number of occasions that they could have been met if perfect care had been provided. The scores therefore represent the percentage of applicable indicators that were successfully met. This same method has been used irrespective of whether the unit of analysis is the condition, modality, or practice.

(b) Investigating the predictors of performance

Having examined summary scores, we then investigated potential predictors of practice performance. These predictors were whether or not practices were GP training practices, their level of computerisation, their level of social deprivation and the quality scores for the different modalities. The outcome variables were the modality quality scores (i.e. summary scores of the condition scores once each condition had been categorised as an acute, preventive or chronic condition). We tested the following specific hypotheses:

- Training practices provide higher quality acute, chronic and preventive care.
- Practices with a higher level of computerisation have higher quality scores for chronic and preventive care.
- Higher quality care (for acute, chronic and preventive care) is provided in practices with less social deprivation.
- Practices with high quality scores for preventive care also have high scores for chronic care.
- Practices with high quality scores for preventive care also have high scores for acute care.
- Practices with high quality scores for acute care also have high scores for chronic care.

(v) ***Ethical issues***

Approval of the Local Research Ethics Committees in the two study areas was obtained. Anonymity at a patient, practitioner and practice level was assured at the time of recruitment. In three practices the researchers were asked to sign confidentiality forms. Any lists of patients' names remained at the practices at all times and these were destroyed 6 months after the completion of data collection. The project team agreed that any concerns that the researchers had in relation to patient care should either be raised with staff at the practice or with the project manager.

Each practice received written feedback detailing their quality scores for the individual conditions and were invited, if they wished, to discuss these with the project team.

(vi) Reliability and Validity

Close working between the two researchers to encourage the development of a common frame of reference and the use of written manuals to provide a detailed coding frame maximised the reliability of the data extraction process. In addition, as noted previously, at the beginning of the fieldwork fifteen sets of records were audited by both researchers and the results compared in terms of the conditions identified and how the records had been coded.

Inter-rater reliability was formally assessed in relation to five conditions (depression, family planning, hypertension, urinary tract infection and tonsillitis). At four practices (two in each of the study PCTs) 25 sets of records for each of these five conditions were audited by both researchers. Items with a Cohen Kappa coefficient of agreement value < 0.60 were excluded from analyses as only values above 0.6 are generally regarded as having an acceptable level of agreement. As a result of this process, data from 10 indicators (6 depression, 2 hypertension and 2 urinary tract Infection) were not used in the analysis.

CHAPTER 5

(i) Results - the indicator set

The evidence base underlying the indicator set, outlined in the detailed literature reviews conducted by primary care experts for each of the conditions, represent an important product of the study and are published in full elsewhere.[36] This chapter presents the indicators for each of the conditions which resulted from the literature reviews and the expert panel process.

Recommended quality indicators for asthma

Diagnosis	
1	1. A diagnosis of asthma should be easily identifiable in the notes 2. Current medication should be recorded in the notes
Management	
2	Written self management plans should be offered to all adults with asthma who: 1. are on high dose inhaled steroids 2. have been hospitalised with asthma
3	Patients with asthma, if on medication, should have their normal peak flow measured on at least one occasion
4	Patients presenting with asthma in the last five years but not on current medication should have their normal peak flow measured on at least one occasion
5	Patients with asthma, if on current medication, should have their predicted peak flow calculated on at least one occasion
6	Patients presenting with asthma in the last five years but not on current medication should have their predicted peak flow calculated on at least one occasion
7	1. Patients with asthma, over the age of 12, should have been asked about their smoking status within the last 5 years 2. Patients with asthma over the age of 10 should have been given smoking advice 3. Smokers should be advised how to stop using a combination of advice and support from a health professional

continued overleaf

8	Patients on current medication or presenting with asthma should have their inhaler technique checked at least once every five years
9	For patients on current medication or presenting with asthma, patients should be asked at every asthma consultation in the last year about 1. any difficulty sleeping due to asthma 2. any asthma symptoms during the day i.e. cough, wheeze 3. whether asthma has interfered with usual daily activities
10	Patients with asthma should not have been prescribed a beta-blocker unless there is justification for doing so
11	Short-acting 82 agonist should only be prescribed on an 'as required' basis
12	Patients using short-acting 62 agonist >1 time a day should be offered prophylactic medication tailored to their individual needs
13	Patients consulting with an acute exacerbation of asthma should have a PEF taken and this should be compared to their normal or predicted PEF
14	In acute situations requiring emergency treatment the following should be assessed and recorded: 1. pulse rate 2. respiratory rate
15	Patients with an exacerbation should be treated with oral corticosteroids by the GP, unless contraindicated or intolerant, if their PEF is <50% of normal / predicted unless they are admitted to hospital
Referral	
16	Patients should be referred to a specialist if they have: 1. occupational asthma 2. been prescribed, or being considered for, oral steroids as maintenance therapy 3. been prescribed or considered for nebulisers in maintenance therapy

Diagnosis	
1	The diagnosis of CAD should be clearly identifiable on the electronic or paper records of all known CAD patients
Treatment	
2	Patients with CAD should be advised at least once to take aspirin at a dose of 75-150mg/day (continued indefinitely) unless they have a contraindication
3	Patients with CAD should have their blood pressure measured and documented at least every 2 years
4	Patients with CAD should have their BP checked at least yearly if they have a systolic blood pressure > 140 mmHg and/or diastolic blood pressure >85 mmHg
5	Patients with CAD and a sustained systolic blood pressure > 160 mmHg or a diastolic blood pressure >100 mmHg should be offered anti-hypertensive medication (as necessary to attain a mean blood pressure of < 140/85)
6	Patients with CAD and hypertension on treatment and a mean systolic blood pressure of >150/90 should be offered a change in therapy (if not changed in the previous six months)

7	Patients with established CAD should have had their blood lipids measured within the last 5 years
8	Patients with established CAD with a total cholesterol level of >5mmol/L should be offered dietary advice or lipid lowering therapy or a change in therapy (if not changed in the prior 6 months)
9	Patients with established CAD should have their smoking status recorded since their disease has been diagnosed
10	Smokers should be given smoking cessation advice at least once since diagnosis
11	Patients with a history of acute MI should be currently prescribed a beta blocking drug indefinitely, unless specific contraindications exist
12	ACE inhibitors should be currently prescribed for all patients for whom there is documented evidence of impaired systolic function and no specific contraindications or intolerance documented in the records
13	Patients with CAD should have their blood sugar measured once since diagnosis

Recommended duality indicators for depression

Screening	
1	Women over 50 with depressive symptoms should have been screened for hypothyroidism in the last three years
2	For any patient presenting with sleep disturbance or fatigue enquiry should be made about other symptoms of depression (for example, depressed mood, markedly diminished interest or pleasure in almost all activities, significant weight loss/gain, psychomotor agitation/retardation, fatigue, feelings of worthlessness (guilt), impaired concentration and recurrent thoughts of death or suicide)
Diagnosis	
3	In the assessment of depression, enquiry should be made about: 1. alcohol use 2. substance misuse 3. current medication
4	The presence or absence of suicidal ideas should be sought out routinely in all patients found to be depressed
5	Patients with suicidal thoughts should be asked if they have specific plans to carry out suicide
Treatment	
6	Patients with a diagnosis of depressive disorder (low mood or lack of interest in usual activities for two weeks plus four of seven other symptoms and impaired functioning) should be offered an effective first line treatment (antidepressant or cognitive-behavioural therapy or problem solving)
Follow-up	
7	Patients with depression prescribed antidepressant drug treatment should be invited for review by a health care professional within four weeks of initiating antidepressant drug treatment
8	Treatment with an antidepressant should be continued for at least four months after recovery from depression
9	Patients with depression should be referred for a specialist opinion where there is evidence of: 1. a possible psychosis

continued overleaf

	2. organic brain disease 3. the patient exhibiting suicidal behaviour 4. serious self-neglect 5. violent behaviour
The elderly	
10	GPs should ask about the presence or absence of symptoms of depression among people aged 65 and over who have been bereaved in the last 12 months
11	GPs should ask about the presence or absence of symptoms of depression among people aged 65 and over who are suffering from: 1. a recent cerebrovascular accident 2. malignancy (except for skin cancer) 3. early dementia 4. Parkinson's Disease 5. Huntington's Disease 6. Chronic pain 7. Multiple unexplained symptoms
12	Antidepressant treatment should be initiated at half the usual starting dose in patients aged 75 and over

Recommended quality indicators for diabetes meilitus

Diagnosis	
1	The diagnosis of diabetes should be clearly identifiable on the electronic or paper records of all known diabetics
Treatment	
2	If the HbA1 c level of a diabetic patient is measured as >8%, the following options should be offered: change in dietary or drug management; or explanation for raised test; or written record that higher target level is acceptable
3	HbA1c levels should be checked in diabetic patients at least every 12 months
4.1	If a diabetic has a sustained blood pressure recorded as >140/85 mmHg on three or more consecutive occasions then a change in non-drug or drug management should be offered
4.2	Diabetic patients with a blood pressure of > 140/85 should have their blood pressure remeasured within three months
5.1	Diabetics should have their feet examined at least once every 12 months
5.2	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
6	All diabetic patients should have an annual fundal examination
7	All diabetic patients should have the following measurements taken for lipid profile within the last 3 years: 1. total serum cholesterol 2. triglycerides
8	Diabetic patients with established Ischaemic Heart Disease and a raised fasting cholesterol (> 5mmol/L) should be advised about dietary modification, or to take lipid lowering medication

9	Diabetic patients with sustained proteinuria should be currently prescribed treatment with ACE inhibitors, unless contraindicated
Follow-up	
10	Patients should be seen by an appropriate health care professional (GP, Practice Nurse, Diabetic Dr) annually
11	All diabetic patients should be offered: 1. influenza vaccination annually 2. pneumococcal vaccination unless contraindicated or intolerant

Recommended quality indicators for hy

Screening	
1.1	All adults over the age of 25 years should have had their blood pressure measured in the last 5 years
1.2	Patients with a blood pressure of > 160/100 should have their blood pressure re-measured within 3 months
Diagnosis	
2	Blood pressure should be measured on at least 3 separate days before starting drug treatment unless blood pressure > 190/140
3	The diagnosis of hypertension should be clearly identifiable on the electronic or paper records of all known hypertensives
4	Initial history should document assessment of the following within three months of diagnosis 1. personal history of peripheral vascular disease 2. diabetes 3. hyperlipidaemia 4. smoking status 5. alcohol consumption
5	Initial laboratory investigations should include the following tests within 3 months of diagnosis: 1. urine strip test for protein 2. serum creatinine and electrolytes 3. blood glucose 4. serum/total cholesterol 5. ECG
Treatment and follow-up	
6	Drug therapies should be offered in all patients with sustained (on more than 3 occasions) systolic BP >/= 160 mm Hg or sustained diastolic BP >/= 100 mm Hg despite up to six months of non-pharmacological measures, unless contraindicated or intolerant
7	Drug treatment is offered in patients with sustained (on more than 3 occasions) systolic BPs of 140-159 mm Hg or diastolic BPs 90-99 mm Hg if despite six months of non-pharmacological measures: 1. target organ damage is present (defined as an abnormal result on any of the tests/exams that pass) 2. there is evidence of established cardiovascular disease 3. the patient is diabetic 4. the 10-year CHD risk is > 30%

continued overleaf

8	All patients with a diagnosis of hypertension should have the following non-pharmacological measures recommended: 1. weight reduction if BMI > 30 2. limitation of alcohol consumption
9	Unless clear contra-indications are recorded, non-diabetic patients should currently be prescribed as first line therapy either a thiazide diuretic or a beta blocker
10	Patients with the conditions below should not be treated with the following drugs: 1. Beta blockers for patients with a history of asthma 2. ACE inhibitors for pregnant women
11	Patients prescribed antihypertensive medication should have their blood pressure recorded at least once per year
12	Patients with sustained high readings (> 150/90 on 3 or more occasions) who are already taking antihypertensive medication should be offered a change in therapy
13	Patients prescribed ACE inhibitors should have had their renal function checked: 1. within the six months before starting treatment 2. one month after the start of treatment

Recommended quality indicators for osteoarthritis

Treatment	
1	Patients with a new diagnosis of osteoarthritis who wish to take medication for joint symptoms should be offered a trial of paracetamol if not already tried
2	If NSAIDs are considered, Ibuprofen should be considered for first line treatment unless contraindicated or intolerant
3	Patients with osteoarthritis prescribed oral NSAIDs who are at high risk of gastrointestinal side effects (past history of dyspepsia or known peptic ulcer) should be considered for a co-prescription of PPIs, H2 antagonists or Misoprolol, unless contraindicated or intolerant
4	Patients with severe symptomatic osteoarthritis of the knee or hip who have failed to respond to conservative therapy should be offered referral to an orthopaedic surgeon for consideration of joint replacement

Recommended quality indicators for acne

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

Recommended quality indicators for acute back pain

Diagnosis	
1	Patients aged 50+ presenting with sudden onset low back pain (onset < 24 hours) should be asked about a history suggestive of spinal fracture (past history of trauma, prolonged steroids, cancer, risk factors of osteoporosis)
2	Patients with referred leg pain (not buttock) should be asked about urinary disturbance
Treatment	
3	X-rays should not be performed in acute lower back pain of less than 6 weeks duration unless "red flag" signs/symptoms exist

Recommended quality indicators for acute diarrhoea in childre

Diagnosis	
1	Children less than 16 years (or their carer) presenting with acute diarrhoea should be asked questions about the following areas: 1. the date of onset or duration of diarrhoea stools 2. presence of blood in stool 3. vomiting
2	Children less than 3 (or their carer) presenting with acute diarrhoea should be asked questions about the following areas: 1. the date of onset or duration of diarrhoea stools 2. presence of blood in stool 3. vomiting 4. fever
3	Children less than 16 years (or their carer) presenting with acute diarrhoea should be asked about their fluid intake
4	Children under 2 years (or their carer) presenting with acute diarrhoea should be asked about urine output
5	Children under 3 years presenting with diarrhoea should be examined with regard to general hydration status
Treatment	
6	Antimicrobial agents should not be used in a child with diarrhoea unless there is a positive microbiological confirmation and the child is not improving

Recommended quality indicators for acute otitis media

Diagnosis	
1	Young children (under 2 years old), presenting in person to the clinician, who have systemic upset (one or more of: fever, irritability, lethargy, vomiting) with no other obvious cause should be examined including an ear examination using an otoscope.
Treatment	
2	Antibiotics should not be offered in children 2 years old and over with uncomplicated acute otitis media (no ENT malformations, recurrent infections, or immunocompromised), unless there is persistent fever, otalgia or discharge 72 hours after seeing the doctor (or 96 hours in total)
3	Children with acute otitis media should not be prescribed oral decongestants

Recommended quality indicators for allergic rhinitis

Treatment	
1	If nasal decongestants are prescribed for patients with allergic rhinitis, then they should not be prescribed for longer than 1 week in any three month period
2	If systemic corticosteroids are prescribed they should: 1. not be for longer than 14 days 2. not be by injection 3. only be prescribed after an adequate course of anti-histamines and topical treatment have proven to be ineffective or were not tolerated

Diagnosis	
1	The diagnosis of peptic ulcer disease should be clearly identifiable on the electronic or paper records
2	For patients consulting with dyspepsia enquiry should be made about: 1. Previous history of peptic ulcer disease 2. UseofNSAIDS 3. Presence or absence of 'alarm symptoms' (weight loss, early satiety, dysphagia, haematemesis, melaena)
Investigation	
3	Patients with 'alarm symptoms' (weight loss, early satiety, dysphagia, haematemesis, melaena) should be referred for urgent endoscopy or specialist referral at first presentation to the GP
4	H.pylori serology should be for initial diagnosis only, not as a test for cure
Treatment	
5	H. pylori eradication should be offered to patients with proven duodenal ulcer disease (confirmatory test not necessary) with active symptoms and who have not had triple therapy within the last month
6	H.pylori eradication regime should consist of a PPI + two antibiotics for a week

Recommended quality indicators for headache

Diagnosis/referral	
1	Patients should be referred urgently for specialist care and investigation if the presenting headache is accompanied by: 1. suspected raised intracranial pressure 2. new onset seizure 3. focal neurological signs 4. papilledema
Treatment	
2	Prophylaxis treatment should be offered in patients with severe and disabling migraine
3	The following agents should be prescribed as first line for prophylaxis of migraine unless contraindicated: 1. beta blocker 2. tricyclic antidepressant 3. pizotifen
4	1. Sumatriptan should not be prescribed for migraine in patients with angina 2. Beta blockers should not be prescribed for migraine in patients with asthma.

Recommended quality indicators for upper respiratory tract infections

Tonsillopharyngitis	
1	Patients with a documented past history of rheumatic fever presenting with tonsillitis or pharyngitis should be advised to take a course of antibiotics unless contraindicated or intolerant
2	If throat infections are treated, treatment should be with penicillin V unless the patient is allergic to penicillin
3	Aspirin should not be prescribed or advised in children with URTIs under the age of 12 years
Bronchitis	
4	Patients with the following symptoms should receive a physical examination of their chest: 1. acute cough with fever persisting for 1 week or deteriorating 2. acute cough with shortness of breath
5	An antibiotic prescription should not be offered to patients with uncomplicated bronchitis with symptoms of less than 14 days duration
Rhinitis	
6	An antibiotic prescription should not be offered to patients with uncomplicated infective rhinitis with symptoms of less than 14 days duration

Recommended quality indicators for urinary tract infection

Diagnosis	
1	In men aged 15+ presenting with dysuria enquiry should be made about a history of urethral discharge
2	Prior to antibiotic treatment, a urine culture should be obtained for patients who have dysuria and any 'complicating' factor (i.e. with complications or where complications are more likely): 1. immunocompromised state 2. suspected diagnosis of pyelonephritis 3. structural/functional anomalies of urinary tract 4. pregnancy 5. men 6. children 7. recent instrumentation of the urinary tract
3	If an infant or child under the age of 12 presents with any of the following symptoms/signs (unless the child is admitted immediately to hospital), a urine culture should be performed : 1. malodorous urine, abnormal urinary stream, or change in urinary stream, or unexplained systemic symptoms (e.g. failure to thrive; jaundice; fever in a neonate) 2. dysuria, frequency, urgency, flank pain (unrelated to trauma) 3. haematuria unrelated to trauma 4. secondary enuresis
Treatment	
4	Patients diagnosed with an upper tract or other 'complicated' UTI should receive treatment with antimicrobials
5	Quinalones should not be used as the first line agents for patients with uncomplicated UTIs without justification
6	If uncomplicated lower tract infections are treated with antibiotics, treatment should not exceed 5 days
7	Patients should be prescribed antimicrobial therapy for at least 7 days for a suspected upper tract infection (Pyelonephritis)
8	Children with suspected or confirmed UTI should be reassessed within 10 days
9	Children less than 5 years old with a first UTI should be referred for specialist opinion within one month.
10	Children aged 5-12 with suspected pyelonephritis who have not had urological investigation, should be referred for specialist opinion

Recommended quality indicators for cervica

Screening	
1	The medical record should contain the date and result of the last smear (for women aged 25-64)
2	Women should be offered routine screening no less frequently than 5 yearly (unless never sexually active with men or have had a hysterectomy for benign indications) unless refusal is documented
3	Women should be offered routine screening no more frequently than 3 yearly (unless the previous smear was anything other than negative or they are immunodeficient)
4	Women aged >65 should not be offered screening unless they have had two abnormal smears in the previous 5 years
Treatment	
5	Women with history of cervical dysplasia or carcinoma in-situ should have had a smear performed within 12 months following the abnormal smear
6	Women with a severely abnormal smear should be referred by the GP for colposcopy within 2 weeks of the receipt of result
7	Women with a low grade lesion should have either a repeat smear or colposcopy within 6 months
8	Women with borderline changes on their smear results who have had the abnormality documented on 3 consecutive smears should be offered referral for colposcopy

Recommended quality indicators for family planning and contraception

Treatment	
1	Women prescribed COCs should be asked about their current smoking status
2	Women over the age 35 who smoke should not be prescribed COCP without justification
3	A woman's blood pressure should be measured when she starts the COCP or have been recorded within the previous 12 months
4	Women prescribed the COCP should have their blood pressure checked within 6 months of starting COC
5	Women with a history of migraine with aura should not be prescribed the COC

Recommended quality indicators for hormone replacement therapy

Screening and diagnosis	
1	Prior to patients starting HRT treatment a doctor or nurse should undertake : 1. a history (including counselling about the risks and benefits) 2. blood pressure check
2	Patients on HRT should be: 1. offered a review including history (including side effects, counselling about duration of treatment) and examination (BP) at least annually 2. encouraged to take part in a national mammography screening programme

continued overleaf

Treatment	
3	Patients suffering from vasomotor symptoms at or after menopause should be offered HRT if other causes are excluded and there are no contraindications
4	Women with early menopause (<45), F.H. osteoporosis, long-term use or repeated courses (>3 times a year) of Prednisolone >7.5 mg/day, early fragility fracture (vertebral fracture without any trauma, fracture neck of femur without major trauma {simple fall is not major trauma}), malabsorption, rheumatoid arthritis should be offered HRT unless there are specific contraindications
5	Women with an intact uterus should not be offered unopposed oestrogen unless the patient has tried and is unable to tolerate oestrogen plus progestogen orTibolone or endometrial sampling is performed periodically.
6	Patients with a history of breast cancer should not be offered HRT except following referral for specialist opinion.
7	Women with a history of deep vein thrombosis or pulmonary embolism, should not be offered HRT unless the risks and benefits have been discussed.

Recommended quality indicators for immunisations

Hepatitis B	
1	Adults and adolescents in the high risk groups should be offered three doses of HBV within 1 year of the following risk factors: 1. Babies of mothers who are chronic carriers of hepatitis B 2. Babies who have had acute hepatitis B during pregnancy 3. Parenteral drug misusers 4. Haemophiliacs or those receiving regular blood products 5. Patients with chronic renal failure on dialysis
Influenza	
2	Adults and adolescents in the following high risk groups should be offered an annual influenza vaccination : 1. Chronic respiratory disease 2. Chronic heart disease 3. Chronic renal failure 4. Diabetes 5. Immunosuppression of any cause 6. Residents of nursing and residential homes 7. Anyone aged over 75
Pneumococcal	
3	Adults and adolescents in the high risk groups except splenectomy should receive pneumococcal vaccination on one occasion: 1. Asplenia or severe splenic dysfunction 2. Chronic respiratory disease 3. Chronic heart disease 4. Chronic renal failure or nephrotic syndrome 5. Immunosuppression of any cause 6. Chronic liver disease 7. Diabetes
4	Adults and adolescents who have no spleen should have received pneumococcal vaccine within the last 10 years

(ii) Results - the demonstration project

This section presents the results of the demonstration project under the following headings:

5.1 Characteristics of participating practices
5.2 Feasibility of data extraction
5.3 Prevalence of conditions and indicators
5.4 Quality of care
5.5 Determinants of quality of care

5.1 Characteristics of participating practices

Most of the participating practices (9/16) had a list size of under 6,000. Half of them (8/16) had between two and four GPs and over two-thirds (11/16) employed between two and four practice nurses (tables 6, 7 and 8).

Table 6: List sizes of the study practices (n=16)

List Size	Frequency	Percentage
1,000-3,000	**2**	12.5
3,001-5,999	**7**	43.8
6,000-8,999	**5**	31.3
9000 plus	**2**	12.5
Total	**16**	100.0

Table 7: Number of general practitioners at the study practices (n=16)

Number of GPs per practice	Frequency	Percentage
1	**3**	18.8
2-4	**8**	50.0
5-7	**3**	18.8
8-10	**0**	0.0
11 plus	**2**	12.5
Total	**16**	100.0

Table 8: Number of practice nurses at the study practices (n=16)

Number	Frequency	Percentage
1	**2**	12.5
2-4	11	68.8
5-7	**2**	12.5
8-10	**1**	6.3
11 plus	**0**	0.0
Total	**16**	100.0

There were equal numbers of training and non-training practices in the sample, reflecting the sample stratification prior to recruitment (table 9).

Table 9: Training status of the study practices

GP Training Practice	Frequency	Percentage
Yes	**8**	**50**
No	**8**	**50**
Total	**16**	100

All the practices apart from one were computerised, the majority using their computer for patient registration, prescribing and basic consultation data (table TO). One practice was 'paperless'. Nationally 97.5% of practices in England are computerised (Simon Richards, personal communication, 2002) but there are no national data available on the level of computerisation.

Table 10: Level of computerisation of the study practices

Level	Frequency	Percentage
No computerization	**1**	6.3
Registration details only	**0**	0.0
Registration and prescribing only	**0**	0.0
Registration, prescribing and some consultation data	**6**	37.5
Registration, prescribing and all consultation data	**8**	50.0
Paperless (consultations, results and letters)	**1**	6.3
Total	**16**	100.0

5.2 Feasibility of data extraction

One of the aims of the demonstration project was to examine the feasibility of data collection using our chosen method. Consistent with other studies conducted by the NPCRDC, we found that this type of quality assessment is both labour intensive and highly dependent upon the quality of both record keeping and the computer systems used by the practices.

The quality of record keeping was found to be variable both in terms of what was recorded and how easy the data were to extract. Handwriting in manual records was sometimes difficult to decipher and data entry onto computer systems was not standardised. Summary sheets in both manual and computerised records were not always updated and were sometimes inaccurate. As a result, the researchers needed to thoroughly examine both the manual and computerised notes to ensure diagnoses and events were correct.

The researchers needed desk space and access to a computer terminal and in some practices this caused practical problems. In most practices the researcher had to work in the reception area, sometimes moving several times in a day. In some practices data collection had to be restricted to odd sessions to fit in with the practice needs. In addition this work can be arduous and both researchers experienced eye-strain during the course of the project due to poor computer interfaces or poor positioning of the computers in some practices. This inevitably reduced the speed of data collection. At the non-computerised practice, the patient records were of high quality. This, coupled with the fact that there was a single source of information on the management of care, actually made data extraction considerably easier.

5.3 Prevalence of conditions and indicators

There was considerable variability in the number of patient records which contained the tracer conditions ('condition prevalence') and for which an indicator could be applied ('indicator prevalence'). Unsurprisingly, the preventive 'conditions' (cervical screening, influenza immunisation and pneumococcal immunisation) were most prevalent (table 11). Allergic rhinitis (0.2%) and hepatitis B immunisation (0.3%) had the lowest prevalence. Some conditions were common, but because of a weak evidence or lack of professional consensus, their indicator set addressed rare elements of care - for example, the headache indicators applied to only 13 of the 232 patients presenting with a headache.

Table 11: Condition prevalence in patient records

Condition	Prevalence (n=1600)
Cervical screening	429 (26.8%)
Influenza immunisation	324 (20.3%)
Pneumococcal immunisation	259(16.3%)
Headache	232(14.5%)
Acute low back pain	212(13.3%)
Urinary Tract Infection	195(12.2%)
Asthma	194(12.1%)
Hypertension	149(9.3%)
URTI - tonsillitis	142(8.9%)
Depression	141 (8.8%)
Dyspepsia/peptic ulcer disease	138(8.6%)

continued overleaf

Osteoarthritis	91 (5.7%)
Family Planning	90 (5.6%)
Acute otitis media	68 (4.3%)
URTI - bronchitis	63 (3.9%)
Acute childhood diarrhoea	59 (3.7%)
Coronary artery disease	53 (3.3%)
HRT	52 (3.3%)
Allergic rhinitis	36 (2.3%)
Diabetes	34(2.1%)
Acne	25(1.6%)
Hepatitis B immunisation	5 (0.3%)
URTI - rhinitis.	3 (0.2%)

In addition there was variation in the prevalence of individual indicators within and across the study conditions. For example, the cervical screening indicators ranged in prevalence from 388 (90.4%, n=429) for indicator 8.2 to 0 for indicator 8.8.

5.4 Quality of care

We examined quality scores at three different levels - for the individual indicators, for the study conditions and for the different modalities (i.e. preventive, acute and chronic conditions) (tables 12, 13 and 14). The key issues relating to summary practice scores and weighting are addressed in Chapter 7.

Across all demonstration practices, preventive care had the highest modality score (61.95%) and acute care the lowest (54.5%). Condition level quality scores above 75% were achieved for 8 of the conditions reported, namely, cervical screening, URTI (rhinitis), headache, otitis media, URTI (tonsillitis), allergic rhinitis, diabetes and coronary artery disease. The conditions with the highest quality scores (100%) were URTI (rhinitis) and acute rhinitis (tables 13 and 14). However, the total number of eligibility events for both of these was very low (2 and 3 respectively). Conditions with low quality scores included hepatitis B immunisation (20%), dyspepsia/peptic ulcer disease (23.2%), pneumococcal immunisation (27.3%) and acne (29.4%) (tables 12, 13 and 14). For hepatitis B and acne these scores might be explained by low sample sizes and total number of eligibility events, but this is not the case for pneumococcal immunisation (n=259, 293 applicable indicators) or dyspepsia/peptic ulcer disease (n=138, 362 applicable indicators).

Table 12: Preventive Care (mean score = 61.95)

[Comprises 6 conditions, 32 indicators, 1159 cases].

Condition	**No. of patients eligible (Condition prevalence) (n=1600)**	**Total no. of eligibility events**	**Condition Score** (%)[1]
Cervical Screening	429 **(26.8%)**	511	85.5
Family Planning	90 **(5.6%)**	229	70.7
Influenza Immunisation	324 **(20.3%)**	409	56.2
HRT	52 **(3.3%)**	131	54.2
Pneumococcal Immunisation	259(16.3%)	293	27.3
Hepatitis B Immunisation	**5 (0.3%)**	**5**	20.0

[1] Number of indicators passed/number of applicable indicators.

Table 13: Acute Care (mean score = 54.5%)

[Comprises 8 conditions, 42 indicators and 974 cases].

Condition	**No. of patients eligible (Condition prevalence) (n=1600)**	**Total no. of eligibility events**	**Condition Score** (%)
URTI - rhinitis	3 (0.2%)	**2**	100.0
Headache	232 (14.5%)	14	92.9
Otitis media	68 (4.3%)	**84**	86.9
URTI- tonsillitis	142 (8.9%)	185	79.5
UTI	195(12.2%)	209	63.2
URTI - bronchitis	63 (3.9%)	**17**	35.3
Acute childhood diarrhoea	59 (3.7%)	309	32.0
Acute low back pain	212(13.3%)	126	31.8

Table 14: Chronic Care (mean score= 60.0%)
[Comprises 9 conditions, 103 indicators and 861 cases].

Condition	No. of patients eligible (Condition prevalence) (n=1600)	Total no. of eligibility events	Condition Score (%)
Allergic Rhinitis	36 (2.3%)	3	100.0
Diabetes	34(2.1%)	344	82.0
CAD	53 (3.3%)	391	75.7
Hypertension	149(9.3%)	829	69.4
Asthma	194(12.1%)	1521	62.0
Depression	141 (8.8%)	815	51.7
Osteoarthritis	91 (5.7%)	75	41.3
Acne	25(1.6%)	17	29.4
Dyspepsia/Peptic Ulcer Disease	138(8.6%)	362	23.2

Examining the individual indicator scores is important (within the limitations of sample size) as it enables areas of both high quality care and underperformance to be specifically identified that overall condition level or modality level scores may disguise. For example, while diabetes has a condition score of 82% (table 14), the indicators scores range from 100% to 47.1%.

5.5 Determinants of quality of care

Examining the individual indicator scores is important (within the limitations of sample size) as it enables areas of both high quality care and underperformance to be specifically identified that overall condition level or modality level scores may disguise. For example, while diabetes has a condition score of 82% (table 14), the indicators scores range from 100% to 47.1%.

5,5 Determinants of quality of care

Further analyses sought to consider potential predictors of good or bad care within the sample of selected practices. However, these findings have to be treated with caution given the size of the sample of practices (n=16), which meant that the study power to find significant correlations was low.

No statistically significant relationship was found between the training status of practices and the quality of care provided in relation to acute conditions (Mann-Whitney U = 34446.500, p= .371) and chronic conditions (Mann-Whitney U = 72875.500, p=.138) but there was a significant association between training status and the quality of preventive care (Mann-Whitney U = 1453330.500, p= .041). Statistically significant relationships were also found, using a Kruskal Wallis non-parametric test, between a practice's level of computerisation and the quality of preventive care (p=0.007) and acute conditions (0.024) but not chronic care (Chi-Square = 2.413; p= .491). In addition, there were no significant associations between quality of care scores and deprivation, nor between quality of care scores across the three modalities. Some of these findings are counter-intuitive and may reflect sampling issues.

CHAPTER 6

Implications for policy and practice

(i) International transfers of quality indicators

The development of clinical quality indicators is an expensive and time-consuming process. For this reason, using indicators developed by another country may speed the time to develop measures and reduce the costs and burdens associated with such development. Based on our research, however, it appears that indicators developed for the health system in one country should not be transferred directly to another country without an intermediate step. Therefore, whilst in principal it makes sense for other countries to utilise the expertise developed in the US, rather than to develop their own indicators de novo, the appropriateness of directly transferring measurement technologies is questionable [39],[40]. The benefits of using the set developed in one country as a starting point, however, include the increased likelihood that a set of equivalent measures will exist in both countries that will allow for international comparisons of performance to be made.

In this section, the details of which have been published elsewhere,[41] we compare the UK and parent US set of indicators and examine the extent to which quality indicators and their associated technologies can be transferred between countries.

Summary of methods

To ensure comparability with the US set for the analysis presented in this section, we used the same cut off points for US and UK sets, which were validity scores of seven or more, and necessity to record scores of four or more, 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[37]). The UK indicators that we finally published[36] were selected on the basis of different cut offs for validity (eight or more) and necessity to record (six or more). 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 (e.g. during admission after acute myocardial infarction), as UK primary care physicians do not normally provide in-patient care. The comparisons reported here are

therefore based on eighteen conditions. For each indicator in the US set, we identified whether there was an exact or near equivalent indicator in the UK set. This analysis requires making qualitative judgements. Examples of indicators which were classified by the UK investigators as 'near equivalents' are shown in box 8. Table 15 shows examples of indicators which were different.

Box 8: 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 therapy should receive oral hypoglycaemic therapy. UK: If the HbA1c level of a diabetic patient is measured as >8%, the following options should be offered six 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 - 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 seven days. Note: Key common point is prevention of rhinitis medicamentosa - difference between 4 and 7 days not of great clinical significance in this context

Table 15: Examples of differences between the indicators in the US and UK sets

Examples of differences between the US and UK indicators	**US indicator (strength of evidence as assessed by US panel)**	**UK indicator**	**Comment**
1: 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
2: 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.
3: Headache	Recurrent moderate or severe tension headache should be treated with a trial of tricyclic antidepressant agents if there is no medical contra-indication (1)	No equivalent indicator	Tricyclics not routlinely used for this purpose in the UK. UK physicians are possibly less interventionalist, and may therefore draw different conclusions when there is only limited trial evidence

4: 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 last five years	Much more frequent screening recommended by US panel
5: Osteo-arthritis	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
6: 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 four Centor criteria, the doctor should discuss the pros and cons of treatment	More aggressive approach to antibiotic treatment by US physicians.

Results

Ninety eight out of one hundred and seventy four (56.3%) of the US indicators had near or exact equivalents in the final UK set (table 16). 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 panels' 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 no 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).

Table 16: 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

continued overleaf

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	

The following section describes in more detail a sample of the conditions 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, compared to 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 7 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 versus 5 years in UK), and lower threshold for colposcopy (2 moderately abnormal smears in the US versus 3 moderately abnormal smears 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, within 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 5 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, 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.*

Conclusion

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, different cut offs for validity and necessity to record were chosen than in the comparative analysis reported in this paper. It is not entirely clear whether this made a substantial difference in the indicator set.

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 (e.g. detailed indicators on the management of hypertension in patients with angina in the coronary artery disease set), and sometimes related to the different health care context (e.g. requirement for registers of patients with angina, diabetes and hypertension in the UK sets alone).

We have focused on differences in professional practice in the results reported here. However, there are a number of other possible explanations for differences between the two sets of indicators which we have not analysed in detail and for which there is little evidence

to judge their importance. For example, the literature reviews were different and there may have been differences relating to the selection of indicators for scoring by the panels, the composition of panels and the conduct of the panel meetings. Finally, the reproducibility of the panel process is not perfect, though the reliability of panels rating the same set of indicators is generally regarded as acceptable. [32-33]

The GAP project 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 United Kingdom and The Netherlands. Nevertheless, there will always be important contextual differences among countries which mean that primary care indicators cannot be transferred from one country to another without going through a process of adaptation.

The results of a primary care quality assessment can be presented at the level of the individual indicator, the condition, the modality, the clinician, the practice or the PCT. There are several important reasons why we might want to move beyond individual indicator scores. First, as we have noted, some indicators have a very low prevalence, so indicator scores drawn from random samples of patient populations are likely to have wide confidence intervals. Second, individual indicator scores present a very partial picture of clinical quality for any given condition or range of conditions. It is not clear the degree to which performance on one indicator is highly correlated with performance on other indicators for the same condition. Finally, the volume of data across a whole set of indicators can be difficult to manage and interpret for clinicians, managers and the public. For these reasons, there has been much interest in producing aggregate or summary quality scores. The following section summarises our experience of aggregation from the demonstration project and from similar work conducted by RAND .

There are a variety of ways of computing summary quality scores and each has its statistical and practical advantages and weaknesses (Personal communication, John Adams, 2001). The method adopted in this study, applying a unitary value to each indicator and condition, has the advantage of being one of the simplest to apply and understand. However, the resulting scores tend to be dominated by high frequency indicators and conditions.

Other methods of computing summary scores are not so influenced by prevalence. For example, to compute an overall quality score for acute care, scores could first be derived separately for each individual acute condition, and the mean of these scores taken as the overall acute quality of care score. This results in each individual condition contributing equally to the overall score, and prevents the most common acute conditions (for example upper respiratory tract infections) from dominating. However, even this method is based on the assumption that all acute conditions are equally important when it comes to measuring the quality of acute care. This method may also mask important differences in performance among the conditions.

An alternative way to create aggregate scores is to assign a different weight to each indicator, prior to combination. Most often such weighting schemes are designed to reflect differences in the clinical importance of the processes included in the quality assessment system. The challenge in doing this for a system as comprehensive as the one we have developed here is that different outcomes are relevant (e.g. birth weight for prenatal care vs. days lost from work for URTI). Quality adjusted life years is one common outcome metric that might be used, but the work required to create weights based on this approach was beyond the scope of the current project.

There are a variety of other approaches that might be taken. An important consideration when selecting a weighting method is the purpose for which the scores will be used. Without further work, it is not possible now to determine the degree to which different methods produce significantly different results and the criteria by which an approach should be selected.

Many of the different methods one might want to adopt for computing summary scores can be expressed mathematically in the form of a weighting system. The following section will focus on some of the issues surrounding such systems.

Using statistical methods to determine the clinical importance of an indicator are usually inappropriate; this is inevitably a subjective judgement which is best made by experts in the field. However, while experts may agree that meeting indicator A, say, is more important for patient health than meeting indicator B, it is more difficult to say how many times more important A is compared to B. This is precisely the type of judgement required for the purpose of assigning weights.

In addition to clinical importance, weights may be used to adjust for the different frequency with which each indicator, or condition, occurs in entities that are being compared. Highly prevalent indicators can dominate summary scores, and it may be desirable to use weights to reduce this effect, irrespective of the clinical importance of the indicators, so that the summary scores reflect the performance of an entity with a standard population. This form of weighting can be used to minimise potentially distorting effects of case-mix differences between practices. For example, two practices, A and B, provide care of equivalent quality on a condition-by-condition basis, but because practice A has a much higher proportion of patients with common conditions requiring standard management, the un-weighted quality score is substantially higher for this practice.

A further question concerns the size of impact that different weighting systems have on the resulting quality scores. If, for example, quality scores remain fairly stable, and the rank order of practices changes little under different weighting systems, then the issue of how to assign weights becomes somewhat academic. We explored this issue using data from the demonstration project and found that the one weighting scheme we tried affected the condition-level scores for individual practices. Table 17 shows weighted and un-weighted quality scores for depression, for the 16 practices in the study (n.b. this analysis was undertaken using a slightly larger set of indicators than was used for the final analysis of depression presented elsewhere in this report, hence the quality scores differ to a small extent). The un-weighted scores represent the overall percentage of applicable indicators

that were met. To produce weighted scores, each indicator was assigned a weight from 1 to 5 based upon prevalence (i.e. the percentage of depressive patients to whom each applied) in descending 20% bands (81-100%, 61-80%, and so on). The weighted scores therefore equalise the impact of each indicator on the final score.

The results demonstrate, firstly, that this approach to weighting generally increases the quality scores: the mean score for practices in PCT 1 has increased by over 8 points, that for PCT 2 by 4 points. This suggests that, on the whole, practices were better at meeting lower prevalence indicators than higher prevalence indicators, hence scores increased after lower prevalence indicators were given more weight. After weighting, five practices saw their quality scores increase by more than 10 points, with the single biggest increase being 15 points (ID 102; from 46.8 to 61.8). One practice saw its score drop by more than 10 points (ID 204). In terms of rank order, on the whole these changed little, with just one practice changing position by more than 4 places, although the practice in question did move from being near the middle to close to the bottom position (ID 203). Whether this approach to weighting would be viewed as the most appropriate will require further study and discussion; certainly alternate methods should be tested and evaluated.

Table 17: An example of weighting quality of care scores (depression)

PCT	Practice ID	Number of applicable indicators	% of indicators met	Rank order	Weighted % of indicators met	Rank order
1	101	54	61.1	**3**	68.2	**4**
	102	79	46.8	**11**	61.8	**7**
	103	**60**	50.0	10	58.7	**9**
	104	63	44.4	13	49.6	12
	105	**57**	54.4	**6**	63.2	**6**
	106	**70**	67.1	**1**	70.7	**2**
	107	45	51.1	**9**	61.2	**8**
2	201	84	45.2	12	46.7	13
	202	48	62.5	**2**	77.2	**1**
	203	39	51.3	**8**	45.3	14
	204	32	40.6	14	28.8	16
	205	**40**	37.5	16	41.9	15
	206	55	58.2	**4**	69.0	**3**
	207	47	57.4	**5**	65.0	**5**
	208	31	38.7	15	51.1	11
	209	37	54.1	**7**	56.6	10
Mean for PCT 1			53.6	-	61.9	-
Mean for PCT 2			49.5	-	53.5	-

The indicators developed in this project represent the most comprehensive set of clinical quality indicators produced for primary care in the UK. While this represents a significant achievement, it is important to reflect also on their limitations. First, for several of the conditions, only a small part of the management of that condition can be addressed by quality indicators, partly because of the poor evidence base for the management of many conditions in primary care. Secondly, it is important to remember that a disease focused approach to measuring quality in primary care will inevitably miss out many important aspects of care, such as issues relating to access and inter-personal care. They therefore encourage a fragmented approach to a holistic and integrated discipline. This is not a criticism of the indicators developed in this project but it is important to draw attention to their limitations as measures of the quality of primary care.

Bearing in mind some of their limitations, how are these indicators likely to be used? As outlined in earlier chapters, there is considerable interest in measuring quality, and therefore a number of uses to which indicators might be put. In the next section of this chapter, we consider how the indicators might be used by practices with a focus on quality improvement, and how they might be used by PCTs or by government, with a focus on performance management. We examine the extent to which these two approaches are compatible, and make specific reference to the possible role of a new GP contract, in terms of providing incentives for quality improvement.

Use of the indicators by general practices

We anticipate that one common use of these indicators will be by clinicians in individual practices, as they increasingly need to assess the quality of care they provide. There are a number of reasons for this, but perhaps two will dominate GPs' thinking: the incentivisation of quality in a new GP contract, and the need to provide evidence for revalidation. So, whereas the motivation for audit or quality assessment was in the past largely internally driven, there is now an increasing need for general practitioners to use quality indicators to demonstrate that they are providing good care.

Practices themselves are likely to use these indicators by choosing a subject which is of personal importance to them. This may be because they have concerns about their care, e.g. because of a recent significant or critical incident. Or they may choose an area that has been suggested as a priority by others, e.g. as part of the clinical governance programme of their primary care group or trust. If suggested from outside, practices are more likely to engage with a topic if they believe the area to be one in which there is significant health care gain to be achieved.

The primary purpose of using indicators in practices is to stimulate discussion about potential areas for quality improvement activities. Practices need to be cautious about drawing definite conclusions about their care on the basis of using indicators. There are a number of reasons for this. First, for almost all the conditions, what we have produced only allows practices to look at part of the care for that particular condition. That may be because our panels have not been able to develop indicators (e.g. for aspects of care that are

difficult to define), or because there are aspects of care which are not normally recorded in the patient's notes, and so are unsuitable for assessment using an approach based on record review.

Since the care of some conditions is only partially covered by the indicators, practices will need to think about what aspects of care are not included and whether those omissions are important. Are there aspects of care that can't be measured, but none the less will benefit from discussion between the doctors and nurses in the practice? There may be important differences in approach of individual clinicians, or issues practices are uncertain about. Using the indicators can lead practices to talk about wider aspects of care.

If the use of quality indicators points to problems in the care which is being given, then practices need to think about how the data for the indicator have been collected. Does the problem relate to inadequate information in the medical record? If so, does that matter? The panels that developed the indicators judged that the information required to assess them should be in the records. Do members of the practice agree? Or is the problem one of computer coding? GP computing systems give scope for substantial variation in the way in which doctors and nurses record diagnoses. Does this matter? How can the practice agree a set of Read codes so that they can do this type of assessment more easily?

Practices may decide that the way they record information in the notes is acceptable, but there appear to be real problems in the care they are providing. This is a judgement. Indicators only 'indicate'. There may be some patients whose care would be inappropriate to assess using particular indicators. This issue could be addressed by allowing 'exception reporting' - i.e. identification of patients who can be excluded from the numerator and denominator when making an assessment of a practice's care. Practices need to form a value judgement about what the indicators mean in the context in which they are being used. Where care does not reach the standard suggested by the indicators, practices will find the literature reviews in our book [36] useful in deciding how important that aspect of care is, and what they should do about changing practice.

Information from quality indicators can be used towards the practice development plan or a doctor's personal learning plan. In the latter case, it will then form part of the doctor's revalidation folder. The results do not have to be excellent to contribute to a revalidation folder: What has to be shown is critical interest in the practice's work, and a willingness to learn or institute change when important deficiencies are found.

Use of the indicators by primary care trusts and by policy makers

Primary care trusts have responsibility for providing high quality care in their locality. Furthermore, there are increased calls for public accountability of care in the NHS. With this background, primary care trusts may be interested to use indicators for quality assurance or performance management. In future, they are also likely to receive information on practices' performance as part of the payment system for general practitioners. In the previous section, we have suggested that practices need to exercise caution in using these indicators for quality improvement in the setting of individual practices. Primary care trusts need to be even more careful in using them to compare different practices, or as a performance management tool.

Policymakers may use these and other similar indicators in a broader performance management context - e.g. contractual payments for general practitioners, and the indicators and literature reviews developed in this project have been used as background information by those negotiating financial incentives in a new GP contract. In future, there are likely to be financial incentives for meeting both clinical and organisational targets, with payments associated with the achievement across a range of chronic conditions. This strategy is likely to be effective in producing at least some changes in professional behaviour.[4243]

Bearing in mind the pitfalls associated with using indicators in practices, how can primary care trusts and policy makers maximise the likelihood that they will be useful in a broader context, and minimise the risk of unexpected negative effects?

There are a number of negative effects of using indicators that need to be considered in this context. The first is that when indicators are 'imposed' from outside the practice, whether by a PCT or as part of a contract, the ownership of the indicators by practices will be small. They may therefore either ignore requests for information, or provide information which is of poor quality. The provision of financial incentives is likely to increase the chance that data will be provided and may provide motivation to improve the accuracy of the data on which performance is based. However, if there are sanctions or financial incentives associated with providing information to PCTs, then the risk of gaming (e.g. collecting the information in a way that portrays the practice in the best possible light) is increased. If the motive for practices to use indicators relates to their own professional values, there is little incentive other than to be open and honest. That may not be the case when indicators are imposed from outside, or are associated with substantial financial incentives.

What can PCTs and policy makers do to reduce these risks? The first and most important is that the managerial agenda (e.g. wishing to monitor performance, administer incentive payments) should be as closely aligned to a professional agenda as possible. Two conditions from the indicator set we have developed will illustrate this. The first is coronary artery disease. Even among those who are suspicious of evidence based medicine, there is a general acceptance of the sort of guidelines (e.g. National Service Framework) that have been used to generate these indicators. GPs may not like the imposition of quality targets, but they are unlikely to dispute that coronary heart disease is an important condition where improving care in general practice may have a significant effect on morbidity and mortality. Acne, by contrast, is a condition which from a medical perspective causes relatively little serious morbidity, and where the indicators relate mainly to avoiding teratogenic drugs in pregnancy. This is a rare problem in the management of acne, and not one to which GPs are likely to want to devote a great deal of time or effort.

So the first and most important principle of using quality indicators in a contract or performance management setting is to use indicators where the potential health gain is great, and there are well founded evidence based management principles for achieving such gain. The managerial and clinical aims are then most likely to be aligned.

Secondly, primary care trusts need to be aware that all the problems of collecting information on individual practices are compounded when information from different

practices is compared. There may, for example, be differences in the way information is recorded in different practices. If computer records are used, doctors may use different codes, and it may be difficult to get comparable information from different computer systems. The circumstances of different practices may also produce difficulties in making comparisons: practices within one primary care trust can vary widely in social mix, so the case mix of patients seen in practices within one primary care trust can vary considerably. The government has invested significant resource into improving the standards of GP computing through the PRIMIS project (http://www.primis.nhs.uk/), but this is unlikely to have reached the practices who have most need of it. One strategy for improving the quality of data collection is to start to use the data with the hope that people will address problems of coding and recording, when they see the results being fed back, for example, in tables of comparative performance.

Some people may feel that all these problems are so great that the path of collecting information on quality of care using indicators is not worth while, and that detailed indicators have, for example, no place in the contract payment system for general practitioners. This is not our view. We accept that the problems, especially those of data quality and comparability, are considerable, especially where information has to be extracted manually from records. We believe that one way to improve available information on quality is to start to use it. If doctors and nurses start to judge themselves on information they believe is unreliable, this acts as a powerful stimulus to change the way in which they work. However, that reinforces the importance of working in areas which are of importance and relevance to the clinicians involved.

Primary care trusts are in a two way bind. In order to get committed involvement in quality improvement activity, they need to work in ways which engage practices as far as possible in the process. PCT managers recognise the substantial challenge in changing the culture of general practice to one which engages with quality improvement.[44] Shekelle identifies four main reasons why doctors don't more enthusiastically embrace quality improvement schemes.[45] First is that they may not agree with the criteria being used. Second they may regard such schemes as a means of blaming them for anything bad that happens to patients. Third, they see quality improvement schemes as often being un-resourced, and therefore 'extra work'. Finally, they find few 'role models' in terms of schemes which have produced major changes in patient care.

PCTs are beginning to be successful at addressing some of these issues. What is most marked is that there has been a widespread involvement of practices at local level in discussing quality improvement issues. For example, more than two thirds of primary care trusts now organise educational events across all the practices in their area - events which often have very high attendance rates, and 96% of practices are now providing PCTs with data on their care for coronary heart disease (unpublished data from the Third national tracker survey of primary care groups and trusts 2002, NPCRDC). These represent a major change compared to only a few years ago when GPs had little or no sense of responsibility for or involvement in the overall health and health care of local populations. In addition, 70% of PCTs are now offering financial incentives for quality improvement (in addition to prescribing incentive schemes). In future, the provision of specific payments to GPs for

meeting quality targets may have a further significant effect in changing the cultural background to quality improvement in general practice.

On the other hand, PCTs are responsible for the quality of care in their geographical area, and need mechanisms to be able to detect and manage poor performance. Our research shows that a considerable tension remains between these two approaches, and that clinical governance leads in particular are managing the tension by, effectively, ignoring their performance management role, and concentrating on engaging practices with quality improvement.[46]

Is this tension between externally imposed indicators and committed engagement by practices unavoidable? Our research suggests that it may not be. In particular NPCRDC's evaluation of the East Kent PRICCE scheme has provided some useful insights into widespread adoption of quality targets by a large group of general practitioners[47] (full report is available atwww.ekentha.sthames.nhs.uk/Framesets/hpro_clinical_governance.asp). In this scheme, practices were invited to join an initiative in which they were paid a modest financial incentive for achieving a wide range of clinical quality targets determined by the health authority. Not only did the large majority of practices become involved in the scheme, but GPs in East Kent described substantial changes in their own attitudes as a result of participation:

> *"I think you can't separate patient care and professionalism, because you feel good about yourself if you deliver good care. At a time when GP morale is low, I see mine as being the highest it has ever been. All you need to do is to look within your own practice and see the benefits, and see the professionalism in your own practice, and get a boost from that. Forget about the politics, but practice medicine." (GP)*

> *"I think it has had an incredibly dynamising effect on practices. And I think it has been a really traumatic experience for some of the practices. Even the practices we would have felt of as being the best, when they found they weren't went away and did something about it. I found their reaction very encouraging... 'Never mind the project - it isn't good enough - this is my patient and I'm going to do better than that' was the attitude." (Health Authority Manager)*

These quotes reinforce the importance of aligning professional and managerial agendas in order to produce engagement by practices with a process of change. However, financial incentives were also important in East Kent, as were including a range of evidence based strategies designed to change professional behaviour, e.g. academic detailing and alignment of local education programmes with the project objectives.

The literature of professional behavioural change is somewhat depressing. Single interventions rarely produce major effects, though multiple interventions are somewhat more effective than single ones.[48] If our experience of change in PRICCE practices and practices who were Personal Medical Services pilots points to any one thing (National Evaluation of first wave NHS Personal Medical Services Pilots 2002, available at http://www.npcrdc.man.ac.uk/Pages/Publications/Pub_intro.htm), it is about arriving at a shared agenda between clinicians and managers and having a clear agenda about the changes that are needed. Application of educational interventions and financial incentives in that situation may then produce significant change.

This project has resulted in the most comprehensive set of clinical indicators that have ever been developed for use in UK general practice. In addition, it has started to field test these indicators so that primary care practitioners and managers can understand how feasible they are to use and their scientific properties. However, and as with most leading-edge projects, we are left with some important questions about the use of indicators and the role of measurement in general practice unanswered. This final section briefly outlines the key issues that need to be addressed in order to maximise the benefits and minimise the risks of introducing greater measurement to British general practice.

The quality of patient records

We have highlighted problems with access to and quality of information in patient records as a major barrier to driving the agenda forward. Some progress is being made in this area, with the implementation of the PRIMIS IM&T educational programme (http://www.primis.nhs.uk/) and the development of automated data extraction software programmes such as MIQUEST and PROFESS. However, without a major investment in information systems in the NHS, quality assessment will always be expensive, time-consuming and of questionable reliability and validity. Only good information technology will enable the improvement agenda to move forward.

Comprehensive assessment of general practice care

We explicitly developed indicators relating to the technical processes of providing clinical care in the community. Whilst this is an important area, it represents only one part of what general practitioners and primary care nurses do. A comprehensive assessment of quality would need to examine the quality of inter-personal care, the relationship between primary care and other services, patient experience of care, the organisation and delivery of services and their cost effectiveness. Some of these dimensions of care have received less attention than medical care to date. Future work in this field should focus on further developing specific measures in these domains and on understanding the relative importance of the domains and how they relate to each other. Measuring these dimensions of care will require different methods, such as patient surveys, site visits to institutions, and analysis of secondary data.

Keeping the indicator set up-to-date

Rapid advances are being made in our knowledge of clinical conditions and this is reflected in advances in the scientific literature and changes in professional opinion. As a result, some established quality indicators need to be updated and new ones will need to be added. This requires a commitment to continuous investment in the process.

Understanding the characteristics of the indicators

Whilst this project describes our preliminary field testing of the indicator set, our knowledge about the properties of the indicators is still at an embryo stage and, as a result, there is a risk that we will make inappropriate judgements about quality when using the indicators. For example, at present, we know little about how to tighten up some of the less reliable indicators, or about how sensitive many of the indicators are to changes in quality. It is important to address issues such as this before we use them to make the sort of

definitive judgements that are required for performance management purposes. However, using these indicators as the basis for selecting targets for quality improvement may provide needed insight into some of these questions.

Maximising public, professional and managerial engagement

Quality indicators are still regarded largely as managerial tools and with suspicion by professionals and disinterest by the public. If their potential is to be realised, this is probably going to have to change. However, at present we know little about what the various stakeholders think about indicators, why they think in this way and how to maximise their engagement. In an ideal world, all clinicians would regard data collection and analysis for audit purposes as a basic professional responsibility, all managers would understand the contribution of measurement to improvement and how to make appropriate judgements using specific measures, and the public would demand information about quality of care and know how to make use of it. We are a long way from this and putting all our resources into developing new measures and understanding the scientific properties of them will not bring us any closer. In-depth qualitative methods have much to contribute to our understanding in this area.

References

1. Marshall, M.N. Time to go public on performance? *British Journal oj General Practice* 1999; 49: 691-692.
2. Brook, R.H., McGlynn, E.A., Cleary, RD. Measuring Quality of Care. *The New England Journal of Medicine* 1996; **335:** 966-970.
3. Smith, R. All changed, changed utterly. [Editorial]. *BMJ* 1998; **316:** 1917-1918.
4. Department of Health, *A first class service, quality in the NHS.* 1998, Department of Health: London.
5. Department of Health, *The NHS Performance Assessment Framework.* 1999, Department of Health: London.
6. Department of Health, *The NHS plan: a plan for investment, a plan for reform.* 2000, Department of Health: London.
7. Royal College of General Practitioners. *The Future General Practitioner: Learning and Teaching.* 1972, London: RCGE
8. Scally, G., Donaldson, L.J. Clinical Governance and the drive for quality improvement in the new NHS in England. *BMJ* 1998; **317:** 61-65.
9. Majeed, E, Voss, S. Performance indicators for general practice Will lead to league tables of performance. BMJ 1995; **311:** 209-210.
10. Roland, M., Marshall, M.N. General practice in an age of measurement. *British Journal of General Practice* 2001: 611-612.
11. McColl, A., Roderick, P., Gabbay J., Smith, H., Moore, M. Performance indicators for primary care groups: an evidence based approach. *BMJ* 1998; **317:** 1354-1360.

12. Campbell, S.M., Roland, M.O., Shekelle, EG., Cantrill, J.A., Buetow, S.A., Cragg, D.K. Development of review criteria for assessing the quality of management of stable angina, adult asthma, and non-insulin dependent diabetes mellitus in general practice. *Quality in Health Care* 1999; 8: 6-15.

13. McColl, A., Roderick, E, Smith, H., Wilkinson, E., Moore, M., Exworthy, M., Gabbay J. Clinical governance in primary care groups: the feasibility of deriving evidence-based performance indicators. *Quality in Health Care* 2000; 9: 90-97.

14. McGlynn, E.A., Asch, S.M. Developing a Clinical Eerformance Measure. *American fournal of Preventive Medicine (RAND Health Program Reprint Service 98-9Q)* 1998; 14: 14-21.

15. McGlynn, E.A., Kerr, E.A., Asch, S.M. New Approach to Assessing Clinical Quality of Care for Women: The QA Tool System. *Women's Health Issues* 1999; 9: 184-192.

16. Campbell, S.M., Roland, M.O., Buetow, S. Defining quality of care. *Social Science and Medicine* 2000; 51: 1611-1625.

17. Kerr, E.A., Asch, S., Hamilton, E.G., McGlynn, E.A., *Quality of Care for General Medical Conditions: A Review of the Literature and Quality Indicators.* 2000, RAND Health: Santa Monica, CA.

18. Kerr, E.A., Asch, S., Hamilton, E.G., McGlynn, E.A., *Quality of Care for Cardiopulmonary Conditions: A Review of the Literature and* Quality *Indicators.* 2000, RAND Health: Santa Monica, CA.

19. McGlynn, E., Damberg, C, Kerr, E., Schuster, M., *Quality of Care for Children and Adolescents.* 2000, RAND Health: Santa Monica, CA.

20. Cantrill, J.A., Sibbald, B., Buetow, S. Indicators of the appropriateness of long term prescribing in general practice in the United Kingdom: consensus development, face and content validity, feasibility, and reliability. *Quality in Health Care* 1998; 7: 130-135.

21. Donabedian, A. *Explorations in quality assessment and monitoring Volume 1: The definition of quality and approaches to its assessment.* 1980, Michigan: Health Administration Fress, Ann Arbor.

22. Buck, D., Godfrey, C, Morgan, A., *Performance indicators and health promotion targets.* 1996, Centre for Health Economics, University of York: York.

23. Lawrence, M., Olesen, E Indicators of quality in health care. *European fournal of General Practice* 1997; 3: 103-108.

24. Campbell, S.M., Braspenning, J., Hutchinson, A., Marshall, M.N. Research methods used in developing and applying quality indicators in primary care. *Quality in Health Care* In Fress.

25. Maxwell, R.J. Dimensions of quality revisited: from thought to action. *Quality in Health Care* 1992; 1: 171-177.

26. Seddon, M.E., Marshall, M.N., Campbell, S.M., Roland, M.O. Systematic review of studies of quality of clinical care in general practice in the United Kingdom, Australia and New Zealand. *Quality in Health Care* 2001; **10:** 152-158.

27. Ovretveit, J. *Health service quality: An introduction to quality methods for health services.* 1992, Oxford: Blackwell Scientific Publications.

28. Braspenning, J., Drijver, R., Schiere, A.M. *Quality indicators for general practice (Kwaliteits - en doelmatigheidsindicatoren voor het handelen in de huisartspraktijk).* 2001, Nijmegen/Utrecht: Centre for Quality of Care Research/Dutch College of General Practitioners.

29. Eccles, M., Clapp, Z., Grimshaw, J., Adams, EC, Higgins, B., Purves, I. North of England evidence based guidelines development project: methods of guideline development project: methods of guideline development. *BMJ* 1996; **312:** 760-762.

30. Brook, R.H., Chassin, M.R., Fink, A., Solomon, D.H., Kosecoff, J., Park, R.E. A method for the detailed assessment of the appropriateness of medical technologies. *Int. J Technol Assess Health Care* 1986; 2: 53-63.

31. Scott, E., Black, N. Appropriateness of cholecystectomy in the United Kingdom. *Gut* 1991; **32:** 1066-1070.

32. Shekelle, EG., Kahan, J.P., Bernstein, S.J., Leape, L.L., Kamberg, C.J., Park, R.E. The Reproducibility of a Method to Identify the Overuse and Underuse of Medical Procedures. *The New England Journal of Medicine* 1998; **338:** 1888-1895.

33. Kassirer, J.E The Quality of Care and The Quality of Measuring it. *The New England Journal of Medicine* 1993; **329:** 1263-1265.

34. Hicks, N.R. Some observations on attempts to measure appropriateness of care. BMJ 1994; **309:** 730-733.

35. Naylor, CD. What is Approriate Care? *The New England Journal of Medicine* 1998; **338:** 1918-1920.

36. Marshall, M., Campbell, S., Hacker, J., Roland, M. *Quality indicators for general practice: A practical guide to clinical quality indicators for primary care health professionals and managers.* 2001, London: Royal Society of Medicine Press.

37. McGlynn, E.A., Kerr, E.A., Damberg, C, Asch, S., *Quality of Care for Women: A* Review *of Selected Clinical Conditions and Quality Indicators.* 2000, RAND Health Program: Santa Monica, CA.

38. McCormick, A., Fleming, D., Charlton, J., *Morbidity Statistics from General Practice. Fourth national study 1991-1992.* 1995, HMSO: London.

39. McKee, M., Figueras, J. Comparing health care systems: how do we know if we can learn from others? Journal *of Health Services Research and Policy* 1997; 2: 122-125.

40. Mulligan, J., Appleby, J., Harrison, A. Measuring the performance of health systems. *British Medical Journal* 2000; **321:** 191-192.

41. Marshall, M., Shekelle, E, McGlynn, E., Campbell, S., Brook, R., Roland, M. Can health care quality indicators be transferred between countries? *Quality and Safety in Health Care* (in press) 2002.

42. Chaix-Couturier, C, Durand-Zaleski, I., Jolly, D., E, D. Effects of financial incentives on medical practice: results from a systematic review of the literature and methodological issues. *International Journal for Quality in Health Care* 2002; **12:** 133-142.

43. Gosden, T., Forland, E, Kristiansen, I., Sutton, M., Leese, B., Giuffrida, A., Sergison, M., Eedersen, L. Impact of payment system on behaviour of primary care physicians: a systematic review. *Journal of Health Services Research and Policy* 2001; 6: 44-55.

44. Marshall, M., Sheaff, R., Rogers, A., Campbell, S., Halhwell, S., Fickard, S., Sibbald, B., Roland, M. A qualitative study of the cultural changes in primary care organisations needed to implement clinical governance. *British Journal of General Practice* (in press) 2002.

45. Shekelle, P. Why don't physicians enthusiastically support quality improvement programmes? Qualify *and Safety in Health Care* 2002; 11:6.

46. Campbell, S., Sheaff, R., Sibbald, B., Marshall, M., Pickard, S., Gask, L., Halliwell, S., Rogers, A., Roland, M. Implementing clinical governance in English primary care groups/trusts: reconciling quality improvement and quality assurance. *Quality and Safety in Health Care* 2002; **11:** 9-14.

47. Spooner, A., Chappie, A., Roland, M. What makes British general practitioners take part in a quality improvement scheme? *Journal of Health Services Research and Policy* 2001; 6: 145-150.

48. Davis, D., Thomson, M., Oxman, A., Haynes, B. Changing physician performance: A systematic review of the effect of continuing medical education strategies. JAMA 1995; **274:** 700-705.