picker INSTITUTE

making patients' views count

What do you think of your doctor?

A REVIEW OF QUESTIONNAIRES FOR GATHERING PATIENTS' FEEDBACK ON THEIR DOCTOR

ALISON CHISHOLM
JANET ASKHAM

PICKER INSTITUTE EUROPE

SEPTEMBER 2006

Picker Institute Europe

The Picker Institute works with patients, professionals and policy makers to promote understanding of the patient's perspective at all levels of healthcare policy and practice. We undertake a unique combination of research, development and policy activities which together work to make patients' views count. There are three key strands to our work:

- Measurement - researching and evaluating patients' experience
- Improvement - leading initiatives that make improvements happen
- Policy - building evidence to inform health policy.

ISBN 1 905945 01 9

978 1 905945 01 6

Published by and available from:

Picker Institute Europe
King's Mead House
Oxpens Road
Oxford OX1 1RX
Tel: 01865 208100
Fax: 01865 208101
Email: info@pickereurope.ac.uk
Website: www.pickereurope.org

Registered charity no 1081688
Company limited by registered guarantee no 3908160

picker INSTITUTE

making patients' views count

CONTENTS

picker INSTITUTE

making patients' views count

ACKNOWLEDGEMENTS

We are grateful to the Picker Institute Inc. for its support of the Patient-Centred Professionalism project. The Picker Institute Inc. is an international non-profit organisation based in Massachusetts, USA, that supports research in the field of patient-centred care.

We also thank the researchers who responded generously to our queries about their questionnaires. Any faults of interpretation are of course our own.

Chris Graham offered valuable advice on assessing the instruments. Members of the Patient-Centred Professionalism research team contributed through comments and suggestions: Angela Coulter, Liz Cairncross, Sir Donald Irvine.

Professor Crispin Jenkinson, Department of Public Health & Primary Care, University of Oxford contributed to the review of the psychometric evidence.

1 Executive summary

The need has never been greater for rigorous assessment of individual doctors' performance against clearly defined standards. Many of the core qualities of a doctor's performance are best judged by patients, and using questionnaires to gather patient feedback can be an effective approach to assessment which presents a low burden to doctors and patients.

This paper reviews a selection of questionnaires designed to gather feedback from patients on individual doctors. It examines how they were developed; their wording; their coverage of key content domains (such as interpersonal skills, communication, and patient engagement and enablement); and the thoroughness with which they have been tested for validity and reliability[i].

Of the ten instruments selected, three questionnaires are strongest overall in terms of content, development and testing: SHEFFPAT (UK), PAR (Canada) and CAHPS 2.0 (USA). But this review identifies several areas of concern. Some issues of importance to patients are not covered at all in any of the questionnaires. We recommend that:

- Questionnaires more attuned to the patient-engagement agendas of today are developed and include a fuller range of questions
- Further research is carried out to investigate the various elements of physicians' technical competence, patients' capacity to comment on it, and how such findings should be interpreted
- Further consideration is given to the development of questionnaires targeted to specific types of condition or specialty as well as those designed to be administered across a broad range of settings
- More work is done to determine the best way to administer patient feedback surveys in clinical settings
- Instructions for implementation are examined further, as it is impossible to tell from most of the published studies what instructions are given.

There are very many weak questionnaires in existence. Those we have reviewed are among the best available. Even so, most fall short of the ideal, a salutary lesson for those organisations thinking of designing their own.

As the Chief Medical Officer for England has recently said in his report on the regulation of doctors, because there will always be some poorly performing doctors it is vital to "recognise the problems early … and deal with them effectively by rigorous, fair assessment …" (CMO 2006). Patients should be given the opportunity to assess doctors with instruments worthy of that task. Only then will patients' views really count.

[i] We believe this review to be accurate at the time of publication, based on publicly available information about the instruments reviewed. Every effort has been made to identify all published information about the development and testing of all the instruments included in this review. Of course, it is possible that development work has been carried out which is not publicly available, or which was produced after this review was published.

2 Background

2.1 Introduction

This paper reviews a selection of questionnaires designed to gather feedback from patients on individual doctors. Its aim is to assess the strengths and weaknesses of such questionnaires and draw conclusions about which (if any) are best suited for their purpose.

What are patient feedback questionnaires?

- A set of questions about patients' views on the performance of a doctor they have recently consulted
- Administered to a sample of that doctor's patients
- Used alongside other evidence about the doctor's performance, such as number of treatments carried out, assessment by colleagues, success rates
- Used by doctor's employers and/or regulatory body to help the doctor to improve his/her performance

Objectives of this review:

1. To set out the key aspects of doctors' performance for inclusion in such a questionnaire.
2. To highlight principles of good practice in questionnaire development and testing.
3. To examine a selection of patient-feedback questionnaires used by regulatory bodies in the UK, USA and Canada, assessing each on content, development and testing.
4. Finally, to ask which of these questionnaires are best for use in assessment processes (for example appraisal, certification or relicensure) of individual doctors, and whether they are adequate for this purpose.

2.2 Assessing doctors: why this is important

The assessment of individual doctors' performance has assumed increasing prominence worldwide in recent years (Violato et al 2003). This is attributed partly to a more consumerist approach to health care (Davies & Ware 1988) and a requirement for doctors to be more accountable to patients and funding agencies (Levine 2002, Epstein & Hundert 2002, Maudsley et al 2000). It has been prompted also by concerns about patient safety and doctors' poor performance and the resulting drive towards improvements in the

quality of patient care (Wilson et al 1999, Institute of Medicine 1994). In the wake of UK scandals such as Bristol, Alder Hey and Shipman – and the resulting damage to public confidence in the medical profession's standards of practice and conduct - the need has never been greater for rigorous assessment of doctors' performance measured against clearly defined standards as the basis for licensure or certification.

What skills do doctors need?

It has always been expected that doctors should be trustworthy and act in the interest of their patients (Hasman et al 2006). But, as Violato et al (2003) observe, thinking in the western world about medical competence has shifted. Interpersonal skills and competence in patient engagement have come to the fore. It is recognised more than ever that communication skills, interpersonal skills, professionalism and a demonstrated ability to improve continuously sit alongside clinical decision-making and medical expertise as components of competence (Frank 2005, ACGME 2005, Levine 2002, Epstein & Hundert 2002, Maudsley et al 2000, Hall et al 1999, Violato et al 1997). Further, patients nowadays also expect clinicians to respect their autonomy, to listen to them, to inform them, to take account of their preferences, to involve them in treatment decisions and to support their efforts in self-care (Coulter & Magee 2003). Professional guidance and codes increasingly emphasise the importance of these aspects of doctors' practice and conduct, for example Good Medical Practice (GMC 2001) and CanMEDS (Frank 2005). A growing body of evidence shows that people who are actively involved in protecting their health and managing their healthcare have better health outcomes (Coulter & Ellins 2006). If they are to fulfil this role effectively, they require help from clinicians who recognise and actively support their contribution and are willing to engage with them as healthcare partners. Doctors need to: know how to guide patients to appropriate sources of information on health and healthcare; provide effective education on health protection and disease prevention; be able to communicate information on risk and probability in a clear, comprehensible manner; determine patients' role preferences and, where appropriate, involve them in treatment decisions; and provide support for self-care and self-management of chronic conditions (Coulter & Ellins, 2006).

Do doctors meet patients' expectations?

While most patients think well of their doctors most of the time, complaints about insensitive communication or failure to provide relevant information are still relatively common (Coulter 2006a). Up to eighty percent of patient complaints to disciplinary bodies are attributable to a breakdown of communication between patients and doctors (BMA Board of Medical Education 2004). Hutchinson et al (1999) carried out a survey[i] (see endnote) which demonstrated that problems with doctors' manner and attitude constituted the more frequent type of problem and the authors point out that complaints from patients often arise as a result of communication problems.

The frequency of communication problems in patients' negative experiences with doctors has also been found in other research. For example, in 1998, more than half of the 770 complaints to the College of Physicians and Surgeons of Alberta related to issues of physician-patient communication (Hall et al 1999). In a recent MORI study most participants in a series of focus groups (Corrado et al 2005) had had negative experiences with individual doctors. These often related to communication skills, which were recognised by both patients and doctors as being important. They included not

being listened to, doctors not taking the time to talk to them, a lack of rapport, insensitivity and, in the case of older patients, being treated "like idiots."

Hall et al (1999), Kurtz et al (1998) and Southgate (1994) all argue that interpersonal skills are required for key elements of clinical practice such as diagnosis, patient involvement and information provision. Greco et al (2002) cite various studies which suggest that doctors' interpersonal skills have an impact on health outcomes, for example reduced blood pressure (Greco et al 2000, McWhinney 1989), pain control (Stewart 1995) and reduced anxiety levels (Kaplan & Ware 1989). These interpersonal skills are far from cosmetic add-ons to technical skills; rather they are an intrinsic part of good clinical practice. However, complaints do not necessarily reflect the whole range of patients' expectations of doctors. In particular technical competence is still the most highly rated quality, even though interpersonal skills may run a close second (Coulter 2005).

Do patients have a role in assessing doctors?

It is increasingly accepted that many of the core qualities of a doctor's performance are best judged by patients, and gathering feedback from patients is gaining credibility as an approach to assessment (Mercer et al 2004, Greenhalgh & Eversley 1999).

Doctors and patients may not always agree about what constitutes quality in doctors' performance (Mercer et al 2004, Jung et al 1997, Weaver et al 1993, Smith & Armstrong 1989), and may not rate doctors in the same way (McKinstry et al 2004, van Dulmen 2003, Wolliscroft & Howell 1994, McLoed & Tamblyn 1994, Goldman 1994, Institute of Medicine 1994, Klessig et al 1989, Merkel 1984, Beihn & Molineux 1979)[ii]. This is hardly surprising, and should not be taken as evidence that patients' judgements are inaccurate. Patients are *better* placed than anyone to judge many aspects of a consultation. Indeed, Arborelius & Bremberg (1992) cite studies which demonstrate that patients make more valid assessments of the doctor-patient relationship in consultations than do 'independent judges'. In one study which compared patients' and experts' judgements of the patient-centredness of a consultation, patients' judgements were the stronger indicator of health outcomes and efficiency of health care, as measured by the number of diagnostic tests and referrals (Stewart, 1995).

Of course to be useful a doctor would need feedback from a selection of patients. Appropriate sample size is a complex issue and will vary from questionnaire to questionnaire, and depends on the level of reliability required. Studies which have calculated the number of patient responses needed in order to get reliable information have found around 25 completed questionnaires are needed[iii](Crossley et al 2005, Violato et al 2003, Webster 1989).

What use can be made of patient assessments of individual doctors?

For assessing healthcare services at a general level (for example within a hospital or local area or for a particular kind of healthcare service) patient feedback surveys are increasingly seen as a key component of quality monitoring (Cleary 1999). In the UK, USA, Canada, Australia, Denmark, Norway, and many other European countries, findings from such surveys are now widely available[iv]. Although these surveys were designed to improve organisational performance and quality assessment at organisation level, rather than to assess individual practitioners' performance, they show how patient feedback can be harnessed on a major scale.

While systems for gathering feedback from patients at the organisation level are well developed, mechanisms for doing so at physician level are less well established. Yet, the use of a questionnaire, routinely collecting feedback from patients could be a cost-effective means to harness patients' views on the performance and practice of individual doctors. The burden placed by them on patients and doctors is low.

Such feedback can be used to improve performance. Well designed questionnaires gather data which allow doctors to identify strengths and weaknesses in their practice and can direct them to areas where improvement is required (Delbanco 1992). Formative assessment using credible sources of feedback is a powerful stimulus to learning (Davies & Howells 2004, Crossley J et al 2002, Ware et al 1978, Ware 1978). Drawing to the attention of doctors issues such as communication skills can be effective in improving the quality of medical practice (Hall et al 1999, Hearnshaw et al 1996).

Recent UK initiatives have aimed to include such questionnaires as part of routine management: the Quality and Outcomes framework of the New GMS contract encourages GP practices to carry out patient surveys at practice level; and the NHS Appraisal Toolkit also suggests the use of patient questionnaires as one source of evidence for the appraisal of doctors' relationships with patients. Patient questionnaires are incorporated into the Alberta College of Physicians and Surgeons' five year relicensure programme. They are one component of the Physician Achievement Review (sitting alongside feedback from physician colleagues and non-physician co-workers), which provides physicians with formative feedback which allows them to identify areas of practice where they could improve.

Patient questionnaires are not widely used for licensure or certification processes to establish fitness to practise. We believe that routine assessment using valid and reliable instruments could help identify doctors who need training or support.

The next section asks what topics need to be included in patient feedback questionnaires.

3 Questionnaire content

If questionnaires are to provide useful feedback from the patient's point of view, they must assess the skills and qualities of doctors which are important to patients and/or which have been demonstrated to improve the quality of patients' care. We have organised these skills and qualities into the following key domains:

- interpersonal skills
- communication of information
- patient engagement and enablement
- overall satisfaction
- technical competence.

These dimensions of patient experience and satisfaction overlap: for example, communication is to some extent an interpersonal skill, and information is an essential component of patient engagement and enablement. But they are conceptually distinct.

It is also important to record some information about the person answering the questionnaire and about the setting in which the questionnaire is used as both may have a bearing on interpretation of the answers. We discuss each of the key domains, and how they might be assessed by patients, in turn.

3.1 Interpersonal skills

Interpersonal skills cover a range of doctor attributes and behaviours – such as courtesy, openness, empathy - which contribute to the rapport between doctor and patient and facilitate an effective consultation. Empathy[v] has been shown to be a key overall theme in patients' definitions of quality of care (Rees-Lewis 1994) and has been demonstrated to enhance the doctor-patient relationship and to improve patient enablement (Mercer et al 2002) and patient and doctor satisfaction in clinical encounters (Roter et al 1997, Suchman et al 1993). Another overall theme of the doctor-patient relationship is trust. Patients who say they trust their doctor are more likely to adhere to treatment than those who do not.

Thus key interpersonal skills include:

- instilling confidence and trust, being open
- taking a holistic approach, showing empathy, considering the patient's personal situation and concerns
- being 'good with people,' caring and courteous, putting the patient at their ease, not patronising, taking them seriously
- being positive and reassuring
- demonstrating respect for the patient, their privacy and dignity

3.2 Communication of information

Communication of information between doctors and patients in both directions is widely considered a crucial element of a good consultation, and features prominently in contemporary professional guidance, such as *Good Medical Practice* (GMC 2001) and *CanMEDs* (Frank 2005). It involves:

- giving clear, understandable explanations about diagnosis and treatment, and when conducting an examination
- eliciting information from patients about their symptoms and concerns, letting them 'tell their story' and asking about their personal life where appropriate
- being thorough in discussion, encouraging and answering questions
- listening carefully and sympathetically

3.3 Patient engagement and enablement

Enablement is a central part of patient autonomy, and one of the key aspects of patient-centredness. There is a body of evidence that enabling patients to cope with their conditions or illness, and engaging them in their care produces improved adherence and health outcomes (Howie et al 1998). Many standards and codes of practice refer to the importance of engaging and/or enabling patients in some way, although it is an aspect of practice which is too often overlooked (Coulter 2006). Engagement and enablement involve:

- helping the patient to understand and cope with their illness or condition, health and treatment
- involving the patient in decisions about care and treatment
- providing advice to enable patients to keep healthy
- providing information to support self-care, for example about how to take medication, when to return for follow-up care, test results etc
- helping the patient to access other sources of information or support, for example written information, helplines, websites
- giving information about risk in a clear and comprehensible manner
- promoting health literacy, helping patients build skills to access and interpret health information.

3.4 Overall satisfaction

Almost all questionnaires include at least one global rating of satisfaction. Most researchers agree that the value of this sort of question for quality improvement or to stimulate reflection or learning, is limited. Howie et al (1998) argue that items about patients' satisfaction with a consultation reflect their expectations more than the outcome or the benefit that resulted from it. However, the inclusion of a global

satisfaction rating item is often justified in that it provides a benchmark measure against which the validity of other items can be tested. However, we would argue that unless these items have been independently validated and shown to relate to the constructs being measured by the instrument, they may not provide a useful benchmark. It can be measured by asking about:

- overall levels of satisfaction
- whether the patient would recommend this doctor to a friend or family member
- whether the patient would choose this doctor in future.

3.5 Technical competence

Whether or not patients are capable of assessing the technical competence of their doctors depends on what aspects of technical competence they are asked to assess. This is complex. Being a technically competent doctor involves: carrying out appropriate screening or diagnostic procedures or ordering appropriate diagnostic aids (such as scans, blood tests or x-rays), diagnosing correctly, acting with as much speed as the condition warrants, carrying out procedures skilfully (such as surgical operations, injections, physical examinations, blood pressure monitoring), knowing when to refer on or gain other medical opinions, instigating appropriate treatments or preventive actions (such as prescribing medication, ordering inoculations, organising support or therapy), and anticipating the likely outcome of treatment, where this is known (Hasman et al 2006).

Not being medically qualified it is difficult for patients to assess the doctor's performance on many of these aspects, or to give any kind of overall assessment of the doctor's competence (Rao et al 2006). However, patients in certain circumstances could say whether:

- expected/thorough screening or diagnostic procedures were carried out
- in physical examinations or procedures levels of pain or discomfort were kept to a minimum
- a plan to manage a chronic condition had been developed with the doctor
- expected preventive procedures had been offered (such as a flu vaccination for patients over sixty-five) (Coulter 2006b).

It could be argued that even if patients are not good judges of some aspects of technical competence, the fact that they think a doctor is not competent is something the latter needs to know. Although this is not an objective measure of doctors' technical competence, it could be indicative.

The problem for feedback questionnaires of the kind we are considering here is that the instruments are designed to be used with a very broad range of patients, even when they are intended for just one specialty. So it would be hard to include questions solely for older people for example, or just for people with one particular kind of condition, or who have needed a physical examination. One of the questionnaires we have reviewed – GPAQ – had included such questions (eg GPs' diagnostic skills) in an earlier version, but removed them from version two since qualitative interviews with patients indicated that the validity of patients' judgement of doctors' technical competence was low (Bower et al 2002a).

In fact several of the questionnaires we have selected for review do include questions about technical competence. We discuss later how well they handle this topic.

3.6 Other topics for inclusion in the questionnaire

Some questionnaires include questions about organisational factors such as accessibility. These are of great significance to patients' experience of care, but should be interpreted with caution in the assessment of individual doctors since they may not always be within their control. Items can include:

- accessibility of appointments including waiting time for appointments, opening hours and getting through to receptionist on the telephone
- satisfaction with the length of time spent with doctor during the consultation
- access to telephone/email consultation
- whether premises and equipment are accessible, clean, in good repair and provide adequate privacy
- whether other staff are capable, helpful, respectful, work well with the doctor
- referrals to specialists
- continuity/co-ordination of care, particularly in primary care.

Some questionnaires ask for information about the patient such as: age and sex, health status, occupation or other characteristics approximating to social class, ethnic group, number of times the patient has seen this doctor before, and frequency of doctor visits. These factors have been shown to affect responses to patient questionnaires. These items may be drawn or adapted from other national surveys such as the national census to allow comparison with wider populations.

If feedback on a number of individual doctors is to be collated, this information may be used to relate assessments of doctors' performance to inequalities in health. They could be used as the basis for corrective weighting where assessments of doctors' performance are to be compared to one another. In reviewing our selected questionnaires we shall ask whether they include the key topics described in this section. We now turn to how these topics ought to be included in a well-designed questionnaire.

4 Questionnaire design

The quality of any questionnaire must be judged against generally accepted criteria. We do not intend to repeat what can be found in any good textbook, but merely to summarise the key rules for (i) questionnaire development, (ii) wording and structuring of the content, and (iii) testing, so that we can review our chosen instruments against these design criteria. We also discuss how questionnaires are administered.

4.1 Questionnaire development

Translating the topics one wants to investigate into a useable questionnaire is often a lengthy, iterative process. A questionnaire should build on the findings of previous studies and/or the development of similar questionnaires; it usually needs some qualitative investigation to determine how the kind of people who will respond to the eventual survey see the issues; a draft questionnaire needs to be tried out for comprehensibility and completeness; a subsequent draft should be piloted to see if the process of distribution, completion and return works and revised, if necessary, on the basis of this.

A patient feedback questionnaire must be grounded in the values, experiences and understanding of patients (Avis et al 1995) so patient input to the questionnaire development process is crucial, either through the results of previous research or through direct involvement of patients in the development of the new questionnaire.

4.2 Wording of the questionnaire

The wording of instructions, questions and response options should be straightforward, unambiguous and use simple language, so that the meaning is clear to all. Their tone should be courteous and non-patronising. This makes people more likely to complete the questionnaire and increases the chance that they will be able to appreciate the purpose of the investigation, understand the items and respond appropriately. Failure to explain in a covering letter or paragraph who is carrying out the survey, for what reason, and who will see completed questionnaires is inexcusable.

Questions should be appropriate to the concept they wish to probe; for example, an item which purports to measure "depth of relationship" by asking people to agree or disagree with statements such as *This doctor knows all about me* appears to evaluate omniscience rather than the normal interpersonal skills that can reasonably be expected of a doctor!

Asking about a recent, specific point in time makes it easier for the respondent to report accurately on their experiences. Where an overall judgement is required a specific time point is less important. But, asking about how the doctor performs "generally", or getting the patient to think about hypothetical situations (for example, see section 5.3 on wording) will not produce answers which accurately reflect the respondent's viewpoint.

4.3 Testing the questionnaire

Before being put into general use, questionnaires should be tested to see whether they measure what they purport to measure, and whether they measure it consistently and reproducibly. This involves tests of validity and reliability. Each questionnaire was assessed to determine whether such tests had been carried out.

Reliability

Reliability refers to the reproducibility and consistency of an instrument. In other words, would repeat measurements made with the same questionnaire under constant conditions give the same result? There are a number of standard measures of reliability, three being particularly relevant to patient questionnaires: internal consistency; assessment of characteristics by factor structure; and physician-level reliability (see endnote for further details of measures[vi]).

Validity

Validity is a measure of how completely an assessment tool measures what it purports to measure. Several aspects of validity can be distinguished. Those of particular relevance to patient questionnaires are: face validity, criterion validity, and construct validity (see endnote for further details[vii]).

4.4 Mode of administration

Most patient questionnaires have been developed to be administered on paper. There are three main established ways of administering paper-based patient questionnaires:

- Questionnaires are given to a consecutive sample of patients immediately after consultation and are completed immediately and returned to the receptionist or in a box in the clinic area.
- Questionnaires are given to a consecutive sample of patients immediately after consultation and are taken away and posted back to the practice or hospital.
- Less commonly, questionnaires are posted to a random sample of patients who are asked to return them by post.

With paper-based questionnaires, the main advantage of completion immediately post-consultation is that response rates tend to be higher, thus probably reducing non-response bias. However, there is some evidence to suggest that questionnaires administered immediately post-consultation produce more positive data. Bower & Roland (2003) found bias in responses to the General Practice Assessment Survey (GPAS) which related to how the questionnaires were administered: scores were higher when administered at the surgery than by post, with the largest bias in the scale which assessed the receptionist. This suggests that concerns about confidentiality when completing a questionnaire may result in socially desirable responses. Alternatively, this

bias may be the result of a relatively short-lived "halo effect" whereby patients feel more satisfied immediately after their consultation than they do a short time afterwards.

Two newer modes of administration may make the survey more accessible to those who do not read and/or write easily: first, "interactive voice response" (IVR), which uses automated telephone technology, whereby the patient is given a toll-free number to call, then completes the questionnaire over the telephone and, second, completion of questionnaires on line via the internet.

There are advantages and disadvantages of each approach, and the evidence surrounding mixed-mode designs is not conclusive. Rodriguez et al (2006) randomly assigned adult patients to complete a brief, validated patient questionnaire by mail, internet, or IVR. Response rates were higher by mail (50.8%) than web (18.4%) or IVR (34.7%). Mail and web produced identical scores for individual physicians, but IVR scores were significantly lower even after adjusting for respondent characteristics. Conrad & Couper (2004) note high item non-response for IVR and raise a number of concerns regarding its use for surveys, although they point out that their disappointing performance may be due to design shortcomings rather than inherent limitations of IVR itself. Link & Mokdad (2005) compared the respondents to a questionnaire administered by web, mail and telephone and found significant variation in their demographic characteristics, which raises concerns about the comparability of the results.

The majority of patient questionnaires, and most of the questionnaires reviewed in this paper, are still paper-based. For this reason the questionnaires in this review were not compared on this aspect, but we return later to this question at the end to make recommendations on how questionnaire administration could be improved.

After describing our selected questionnaires we review them against the design criteria of development, wording and testing, as well as against their content, as described in section 2.

5 Selection of questionnaires

We selected patient questionnaires used by bodies which have some regulatory function for the medical profession – whether through licensure, certification, accrediting or insuring. Patient questionnaires were identified by searching the websites of the regulatory bodies in the UK, USA and Canada. Where there was no indication that particular patient questionnaires were used, the regulatory bodies were contacted and asked whether they use, or recommend the use of, any patient questionnaires as part of their assessment processes for appraisal, recertification or licensure. Literature was searched for studies of the development and testing of the questionnaires[viii]. Questionnaires were included in this review if they satisfied the following criteria[ix]:

- Designed to gather feedback from patients on individual doctors
- Used by regulatory (licensing, certifying, insuring) bodies in the UK, USA or Canada
- The questionnaires themselves and/or papers or reports detailing their development or implementation were available in the published literature.

We found ten questionnaires meeting these criteria. Table 1 gives basic details.

Genesis, aims and administration of each of the questionnaires

GPAQ was developed by the National Primary Care Research and Development Centre (NPCRDC) in 2003 for use in primary care. It was based on the Primary Care Assessment Survey (PCAS), developed in the USA by Safran et al (1998) and GPAS (Ramsey et al 2000) (see endnote for further details[x]). GPAQ is used for gathering feedback about GPs and to measure aspects of quality of primary care under the new General Medical Services (GMS) contract in the UK. It can be administered by post or immediately after consultation.

DISQ was developed by Greco et al (1995, 1998) for use in primary care, but it has subsequently been adapted for use by the Royal College of Obstetricians and Gynaecologists, and with nurses and hospital doctors. It aims to assess the quality of doctors' interpersonal skills within the consultation (Greco et al 2000). It was designed for use post-consultation, to be returned by the patient to the practice and has been used as a summative and formative tool with GPs for assessment, continuing medical education (CME) and quality assurance in UK, Ireland, Greece, Denmark, Australia and New Zealand (Greco & Pocklington 2001, Greco et al 2001a, Greco et al 2001b, Brownlea et al 1999)

CSQ The Consultation Satisfaction Questionnaire (CSQ) was developed by Baker (1990), Baker & Whitfield (1992) and Poulton (1996) for use in general practice and is recommended for use in the NHS appraisal toolkit. It is designed to be completed by the patient immediately post-consultation and aims to measure patients' satisfaction with their consultation with a GP.

Table one: questionnaires included in this review

Instrument	Developed by	Country	Used for
General Practice Assessment Questionnaire (GPAQ)	National Primary Care Research and Development Centre; Ramsay et al 2000a	UK	New General Medical Services (nGMS) contract
Doctors' Interpersonal Skills Questionnaire (DISQ)	Greco et al 1998; Greco et al 1995	UK	nGMS contract, NHS appraisal toolkit, Royal College of Obstetricians and Gynaecologists (RCOG)
Consultation Satisfaction Questionnaire (CSQ)	Baker & Whitfield 1992a; Baker 1990	UK	NHS appraisal toolkit
CARE Measure	Mercer et al 2005; Mercer et al 2004	UK	Accredited by Scottish Executive/NHS Education in conjunction Royal College of General Practitioners (Scotland) for appraisal of GPs in Scotland, and included in 'toolkit.'
SHEFFPAT	Crossley & Davies 2005; Crossley et al 2005	UK	National Clinical Assessment Agency Royal College of Paediatrics and Child Health Consultant appraisal
ABIM patient questionnaire	Lipner et al 2002	USA	American Board of Internal Medicine (ABIM)
Physician Achievement Review (PAR) Program questionnaires	Hall et al 1999	Canada	College of Physicians of Alberta 5 versions developed for general practitioner, medical specialist, surgeon, anaestheologist and for episodic care
Patient Satisfaction Questionnaire: (PSQ-18) – Short Form; and PSQ III	Marshall & Hays 1994; Marshall et al 1993; RAND Corporation	USA	Planning, administration and evaluation of health services delivery programs, eg. RAND's health insurance experiment
CAHPS 2.0	Hargraves et al 2003; Agency for Healthcare Research and Quality (AHRQ)	USA	Used in USA by health plans, major employers, purchasing groups, and the Centers for Medicare and Medicaid Services
Patient Enablement Instrument (PEI)	Howie et al 1998	UK	This is not used in its own right but has been adapted in GPAQ and other patient questionnaires

CARE The Consultation and Relational Empathy (CARE) Measure was developed between 2001 and 2004 by Mercer and colleagues (Mercer et al, 2004, 2005). It was designed as a patient-assessed measure of communication and empathy for use in general practice and has been piloted successfully in a secondary care setting.

SHEFFPAT was developed by Crossley and Davies (2005) and Crossley et al (2005) initially for use in paediatrics, but has also been used in an adult setting for use in hospital and general practice. Its performance has not yet been formally evaluated for use with adults. It is designed to be given to patients immediately post-consultation by clinic clerk/receptionist, and to be returned to a box in the reception area. Adults are invited to complete the questionnaire on behalf of child patients. Its aim is to measure consultation quality, and it is intended to be used as an assessment tool and as a stimulus for learning.

ABIMQ The American Board of Internal Medicine (ABIM) questionnaire (Lipner et al 2002) is based on an instrument developed by Weaver et al (1993) and Webster (1989). It is used by ABIM to assess aspects of clinical competence not covered by other means, such as humanistic qualities, professionalism and communication skills. It is administered by telephone: the questionnaire is given to the patient by the doctor, post-consultation, and the patient is invited to call a toll-free number and complete it, using an interactive voice response (IVR) system. Its aim is to provide formative feedback for use in Continuing Professional Development (CPD), to encourage and facilitate improvement in practice.

PAR Violato et al (1997) devised patient questionnaires for use with the Physician Achievement Review (PAR) Programme. The PAR programme was set up as a formative performance assessment to provide a multidimensional view of performance through structured feedback to physicians from the physician themselves, their peers, patients and colleagues. It is an element of the Alberta College of Physicians and Surgeons' relicensure process. The programme routinely assesses the performance of physicians, drawing particular attention to physician-patient communication, and is intended primarily to improve the quality of medical practice, but also to identify physicians for whom more detailed assessment of practice performance of medical competence may be needed. Patient questionnaires were adapted for the following specialties: primary care, surgery, anaesthesiology, medical specialties, episodic care. They are designed to be administered on paper immediately post-consultation to 25 consecutive patients.

PSQ-18 The PSQ-18 short form was developed by Marshall & Hays from the RAND Corporation in 1994. It is a shortened version of PSQ III which is, in turn, a shortened version of PSQ, developed originally by Ware et al (1976a, 1976b). This original instrument aimed to "*yield reliable and valid measures of concepts that had both theoretical and practical importance to the planning, administration and evaluation of health services delivery programs"* (Ware et al 1983). The aims of the PSQ-18 are similar but it was developed for use in situations where brevity precludes administration of the full-length PSQ III (Marshall and Hays, RD 1994). It is easier to complete and therefore encourages using patient feedback for monitoring the delivery of medical care. It is designed to gather feedback on medical services as a whole rather than just on individual doctors.

CAHPS 2.0 CAHPS is a family of surveys developed by the Agency for Healthcare Research and Quality (AHRQ) from 1995 onwards. They are widely used survey instruments that ask consumers about experiences with and evaluations of ambulatory care received from health care professionals and health plans (Hargraves et al 2003). Here we focus on the CAHPS 2.0 Adult Core Survey. It is designed for administration by post or by telephone.

CAHPS surveys are now used in the U.S. by many health plans, major employers, purchasing groups, and the Centers for Medicare and Medicaid Services (CMS). It is a component of the accreditation process for health plans administered by the National Committee for Quality Assurance (NCQA) and of NCQA's Health Plan Employers Data and Information Set (HEDIS)[xi]. CAHPS surveys are designed to evaluate how health plans (in the USA) compare with one another, so plan-level (rather than physician-level) reliability is considered key (Hargraves et al 2003).

PEI The Patient Enablement Instrument (PEI) was published by Howie and colleagues in 1998. It was developed for use in primary care and designed to be completed by the patient immediately after consultation. Its stated aim is to capture the concept of 'enablement' which the authors believe reflects patients' ability to understand and cope with their illness. It is not used in its own right by any regulatory bodies, but it is included in this review because elements of it have been incorporated into other questionnaires which are used by regulatory bodies, for example GPAQ.

6 Review of questionnaires

In comparing the ten questionnaires we ask five key questions:

i) Does the questionnaire contain at least some questions on the three key **content** domains of interpersonal skills, communication of information and patient engagement, as well as on overall satisfaction and technical competence?

ii) Were the **views of patients** taken into account in developing the questionnaire?

iii) Was there a thorough **process of development** of the questionnaire?

iv) Is the **wording** of the questionnaire generally clear and straightforward?

v) Were **tests** of validity and reliability carried out?

Of course these are crude questions and the answer may be hard to determine in some cases. It is also possible that we might assess as not fully meeting our criteria a questionnaire which in other respects may be a very useful instrument in accordance with its own aims: we do not intend that our assessment should be taken either as an overall damning condemnation or seal of approval.

6.1 Content

Details of the topics covered in all questionnaires are included in Appendix 1[xii]. Here we summarise the questionnaires' coverage of the three key content domains of interpersonal skills, communication of information and patient engagement, as well as on overall satisfaction and technical competence.

Table two: whether key content domains are included in questionnaire

Questionnaire	Interpersonal skills	Communication of information	Patient engagement & enablement	Overall satisfaction	Technical competence
GPAQ	✓	✓	✓	✓	×
DISQ	✓	✓	×	✓	✓
CSQ	✓	✓	✓	✓	✓
CARE system	✓	✓	✓	×	×
SHEFFPAT	✓	✓	✓	✓	✓
ABIMQ	✓	✓	✓	✓	×
PAR	✓	✓	✓	✓	✓
PSQ-18	✓	✓	×	✓	✓
CAHPS 2.0	✓	✓	✓	✓	✓
PEI	×	×	✓	×	×

Four of the questionnaires include at least one question within all five of our topic categories: technical competence, interpersonal skills, communication of information, patient engagement/enablement, and overall satisfaction: CSQ, SHEFFPAT, PAR and CAHPS 2.0. One other (ABIMQ) includes all except technical competence. For example, CSQ contains 18 items, asking about the most recent consultation, such as general satisfaction, professional care, depth of relationship, perceived time spent with doctor. Technical competence is included through questions on thoroughness of physical examination. It also includes a measure of patient enablement, by asking patients whether they agree or disagree that:

> *I understand my illness better after seeing this doctor*

And it examines patients' experience of whether the doctor took a holistic approach to their care:

> *This doctor was interested in me as a person and not just my illness.*

The SHEFFPAT questionnaire is made up of 13 items plus patient characteristics and a space for other comments. It asks about consultation quality; opportunity to shape discussion; quality of explanation and advice; patient's understanding of condition and treatment; confidence in self-care; interpersonal skills; the doctor's interest in the patient's point of view; respect; confidentiality; overall satisfaction[xiii].

The precise content of the PAR programme questionnaires varies depending on the specialism with which the patient questionnaires are designed to be used. The general practitioner questionnaire covers patient interaction; phone communication; personal communication; information for patients; office staff; physical office; and appointments[xiv].

CAHPS 2.0 has 19 core items, including two global ratings of care, two questions on preventive care, and eight questions assessing communication. Technical competence is assessed through an overall rating of quality of 'medical care', thoroughness and level of discomfort of any examination.

The ABIMQ has 18 items covering issues such as truthfulness; manner; giving information during physical examination; involving patient in decision-making; giving information about problems; giving clear explanations; and not being patronising. For example:

> *How is this doctor at…*
>
> *Treating you like you're on the same level, never "talking down" to you or treating you like a child?*

ABIMQ also recognises the importance of a doctor's capacity to elicit, listen to and understand information from patients as well as to give information. Thus it asks:

> *How is this doctor at…*
>
> *Letting you tell your story; listening carefully; asking thoughtful questions; not interrupting you while you're talking?*
>
> *Showing interest in you as a person; not acting bored or ignoring what you have to say?*
>
> *Encouraging you to ask questions; answering them clearly; never avoiding your questions or lecturing you?*

The other five questionnaires, whose coverage is less complete, vary in the number of items they contain, some understandably being considerably shorter than those which included all domains[xv]. Four have no questions on technical competence; two have no questions on interpersonal skills, two have none on patient engagement, two have no overall satisfaction questions, and one has nothing on communication of information. However, it has to be remembered that different questionnaires have different objectives, which helps to explain the variation in their content. Thus, for example, PEI only contains six questions, all on 'patient enablement' because that is its purpose. It also needs to be remembered that inclusion of questions is no guarantee of quality. For example, some questions on technical competence appear to tap areas of expertise which patients may well not possess (such as '*The medical staff that treats me knows about the latest medical developments*'). We discuss these aspects further in the section on wording.

Perhaps the most interesting finding on questionnaire content is that although there is a good coverage of topic categories in at least half the questionnaires, some *specific* topics do not feature in *any* of the questionnaires (See Appendix 1). These comprise aspects of patient engagement increasingly seen as part of a doctor's role with patients:

- helping patients to improve their health literacy
- communicating risk

Some aspects of doctors' technical competence, such as a plan for managing a chronic condition or offering expected preventive measures also do not feature in any of the questionnaires. But this is not a shortcoming in questionnaires designed to be used with a broad range of patients.

6.2 Questionnaire development

Table three sets out the key aspects of questionnaire development, which are relatively easy to summarise in tabular form.

Table three: Stages included in questionnaire development

Questionnaire	Literature Review	Interviews with patients	Pilot survey
GPAQ	✓	✓	✓
DISQ	×	✓	✓
CSQ	✓	×	✓
CARE system	✓	✓	✓
SHEFFPAT	✓	✓	✓
ABIMQ[xvi]	×	×	×
PAR	✓	✓	✓
PSQ-18	✓	✓	✓
CAHPS 2.0	✓	✓	✓
PEI	✓	✓	✓

On the face of it, most of the questionnaires appear to have been developed according to acknowledged processes, with seven having explicitly involved a review of previous research, some interviews with patients (whether in-depth or cognitive one-to-one interviews or focus groups), and then a pilot of the draft questionnaire. However, although the review of previous research and the piloting of a draft questionnaire are fairly routine matters, the involvement of patients is more complex and interesting. The meaning of 'patient involvement' and how far it should go is a matter for debate. Patients should be asked at an early stage on which areas of performance they feel confident in assessing their doctor, but this rarely happens. In some questionnaires, patient involvement was much more limited than others.

For example, patient involvement in the development of CSQ appears to have been very limited. Issues thought to be of concern to patients were identified from a review of other patient questionnaires and studies in general practice that included surveys of patient opinions. This review was supplemented by discussions with other GPs and from the personal experience of the author (a GP) about patients' comments on their care. Patients' only opportunity to contribute directly to the content of the questionnaire was through two open questions included in the pilot version in which patients were invited to say whether there was anything they particularly liked or disliked about their doctor. No new issues were identified from this. An early draft of the questionnaire was refined first by obtaining colleagues' comments on the meaning of each item and, second, by observing the dispersion of scores on each item. Ambiguities or difficulties in answering questions were detected by studying completed questionnaires and measuring item non-response. These are all legitimate approaches to identifying problems with the questionnaire, but do not encourage active, detailed input from patients and so are insufficiently thorough to ensure the instrument is acceptable to patients.

Much more usual in this selection of questionnaires is a development process which involves patients through interviewing them about whether they understand the questions or response options in a draft questionnaire in the way its authors intended, whether the wording is understandable and acceptable and whether the response options are comprehensive. For example, CAHPS surveys are developed through literature review, interviews with patients as well as other stakeholders (eg managers), and cognitive interviews. CAHPS 2.0 was modified from CAHPS 1.0 on the basis of field studies, additional cognitive interviews and testing with customers to explore their responses to it and to pick up any difficulties or misunderstandings in completing it. On the basis of the testing, some items were refined, dropped or added (Hargraves et al 2003). Similarly the initial development of DISQ was based on patient focus groups which sought patients' views on what constitutes a good consultation and what are the key communication skills in a medical consultation (Greco et al 1995, 1999, 2000). The initial DISQ instrument was then piloted in 30 general practices.

Face and content validity of the CARE Measure were developed through a thorough, iterative process of interviews and revisions with a range of patients, who endorsed the clarity and importance of the items in the final instrument. Tested in both high and low deprivation settings, 76% of patients and 78% of doctors rated the items in the instruments as 'very important' to their current consultation (Mercer et al 2005). Mercer et al (2005) asked patients to rate the relevance of the items in the CARE measure to them. They found that 76% of patients rated the measure as being 'very important' to their current consultation. Higher ratings of importance were observed in older patients, patients consulting with psychosocial problems, patients with long-standing illness or

disability, and patients with significant emotional distress. They report that few patients rated individual CARE items as being 'not applicable' to their current consultation; only 3.1% of patients felt that more than two of the 10 items in the measure did not apply to their current consultation.

Our final example of good practice in questionnaire design is the SHEFFPAT instrument which was originally developed through a review of literature on consulting, a consensus exercise with a group of (paediatric) consultants and the results of a MORI poll (for GMC) on what patients thought made a good consultation. The instruments were then tested in eight iterative pilot trials to ensure that the questions, rating scales and layout were comprehensive, comprehensible and acceptable to raters (Crossley et al 2005).

6.3 Wording

The issue of how topics are converted into appropriate questions and then worded in a comprehensible manner is harder to summarise. Most questionnaires will probably include at least one question which can be criticised for lack of clarity. However, we decided to classify as not meeting our criteria those where (a) there were several examples of unclear wording or unrealistic questions, or (b) several inappropriate response options were used or appropriate response options were omitted. These factors make a questionnaire difficult to complete and affect its reliability and validity.

Table four: Wording of questions and response options

Questionnaire	Clear, appropriate wording throughout	Appropriate response options
GPAQ	✓	×
DISQ	✓	×
CSQ	×	✓
CARE Measure	✓	✓
SHEFFPAT	✓	✓
ABIMQ	×	✓
PAR	✓	✓
PSQ-18	×	✓
CAHPS 2.0	✓	✓
PEI	✓	✓

On the wording of topics, most questionnaires have been sufficiently carefully developed so that their questions are sensible and appropriate. Only three were problematic. In CSQ there were ambiguous items such as ….

> *This doctor knows all about me*
>
> *I felt this doctor really knew what I was thinking*
>
> *There are some things this doctor does not know about me.*

In general, questionnaire items should tap single concepts (Dillman 2002). In ABIMQ some questions are long, complex and refer to multiple issues. For instance:

> *How is this doctor at.. greeting you warmly, calling you by the name you prefer, being friendly, never crabby or rude?*
>
> *How is this doctor at.. warning you during a physical exam about what he/she is going to do and why; telling you about what he/she finds?*

How is a person to answer if for example the doctor does call them by the name they prefer but is sometimes crabby? Or what if the doctor tells them what she finds after an examination but does not tell them in advance what she is going to do and why?

And in PSQ18 some of the items designed to measure technical care are ambiguously worded. It is unclear, for example, whether some of the following questions aim to measure the doctor's interpersonal skill of winning the patient's confidence or directly to measure his/her technical competence.

> *Sometimes doctors make me wonder if their diagnosis is correct…*
>
> *I have some doubts about the ability of the doctors who treat me.*
>
> *I think my doctor's office has everything needed to provide complete medical care…*
>
> *When I go for medical care they are careful to check everything when treating and examining me….*

As we have said above, technical competence is the most difficult topic to operationalise, and although we applaud PSQ18's attempt these questions do not inspire confidence. By comparison the following questions used by CAHPS2.0 are much more specific:

> *In the last twelve months did this doctor ever examine you?*
>
> *IF 'YES':*
>
> *How often did this doctor examine you in a way that caused you as little pain and discomfort as possible?*

Other examples of good practice in question wording were those which asked both about patients' experience and about how satisfied they were with this experience. For instance GPAQ includes some combinations of report and satisfaction questions, such as *In general, how often do you see your regular doctor* and *how do you rate this?* Report-satisfaction pairs are valuable as guides to how practice could be improved because they record not only the patient's experience but allow a comparison of this with the patient's expectations.

As some quite complex and abstract issues are being investigated in these instruments, it may also be good practice to give examples of what is meant by a question. For example, feedback from interviews with patients following completion of an early version of CARE indicated that they would find it helpful to be given more concrete examples of the sometimes abstract items to clarify what they meant. For example,

> *How was the doctor at… making you feel at ease..(being friendly and warm towards you, treating you with respect; not cold or abrupt.)*

On the other hand there is a danger that giving multiple examples may confuse rather than clarify, so this needs to be handled carefully and involve thorough testing with patients.

It is good practice to refer to a time point as specific as possible in questionnaire items. PSQ-18 specifies no temporal reference point, but instructs:

> *If you have not received care recently, think about what you would expect if you needed care today.*

This is very poor practice as it confuses experience with expectation.

On the issue of response options most questionnaires perform well. With fixed choice questions (as is typical for these questionnaires) the range of answers available must include all possibilities, such as allowing respondents to say that they 'don't know', or are 'unsure' about their feelings[xvii]. Only two instruments were weak in this regard: in the DISQ and GPAQ questionnaires *don't know* and *not applicable* response categories are missing for all or some items.

6.4 Tests of validity and reliability

Few survey questionnaires will carry out the whole range of tests of reliability and validity, but a good questionnaire will only gain credibility if certain tests are performed. The validity and reliability of the data gathered by each instrument was determined by assessing evidence for the measurement characteristics of each. This process was carried out by the first author, together with an external expert in psychometrics by examining published reports which documented the development and testing of each questionnaire. A pro forma was designed to ensure that the two assessors judged the instruments against the same criteria: internal consistency; factor structure (where appropriate); physician-level reliability; criterion validity; construct validity. (The face validity of each instrument is covered in section 5.3).

These reports were identified by two means: For some instruments, the validation studies were listed in a single place. GPAQ, for example, has a 'manual' (available on the GPAQ website) which documents the publications which relate to the process of development of the questionnaire. Some publications indicated that further data were available from their authors, in which case this was requested. Development studies for other instruments were identified by searching Web of Science electronic databases by instrument name and author. Reference lists of these papers were also searched until the authors were satisfied that all relevant papers had been identified. Of course, absence of published evidence

does not necessarily imply that testing was not carried out, but without such evidence it is impossible to be confident that the instruments are valid and reliable.

Table five summarises the presence or absence of rigorous evidence for each questionnaire's measurement characteristics. □ denotes either that this aspect of validity or reliability was not tested, or that the testing carried out was judged not to be sufficiently rigorous. Details of the grounds for these judgements are set out in Appendix 2.

Table five: Evidence of rigorous testing for reliability and validity

	Internal consistency	Factor structure	Physician-level reliability	Criterion validity	Construct validity
GPAQ	✓	✓	×	×	×
DISQ	✓	✓	×	✓	✓
CSQ	✓	✓	×	×	✓
CARE	✓	×	✓	✓	×
SHEFFPAT	✓	✓	✓	✓	✓
ABIMQ	×	×	×	×	×
PAR	✓	✓	✓	×	✓
PSQ-18	✓	✓	×	×	×
CAHPS 2.0	✓	✓	✓	✓	×
PEI	✓	×	×	✓	×

As this table shows, instruments' reliability has been tested more often than their validity. The evidence available suggests that only SHEFFPAT rigorously carried out all five key validity and reliability tests, although it has not been fully tested with adult patients. CAHPS 2.0, DISQ and PAR rigorously carried out four of the five.

7 Which of these questionnaires is the 'best buy'?

We are now in position to sum up. Few people would disagree that patients' views about their doctors are important and relevant to doctors' training, continuous assessment and ultimately their fitness to practise. Well-designed and appropriate questions given to a sample of patients by means of a standardised questionnaire are an effective and efficient way to obtain such views.

But are the existing questionnaires fit for this task? Clearly some are better than others, with fuller coverage of the kinds of issue that matter to patients, fewer design faults and more rigorously tested. Appendix 3 summarises the key strengths and limitations of each.

Of the ten instruments selected, those with the fullest coverage of the key issues of concern to patients were CARE, SHEFFPAT, ABIM, PAR and CAHPS. Those most rigorously tested for validity and reliability were DISQ, SHEFFPAT, PAR and CAHPS. Three questionnaires are strongest overall in terms of content, development and testing. Interestingly, one was developed in each of the three countries included in our review:

SHEFFPAT (UK)

PAR (CANADA)

CAHPS 2.0 (USA).

The three questionnaires reflect good development processes and are well worded.

But is this the end of the story? Clearly the answer is 'no', because this review has identified several areas of concern, particularly that:

- Some issues are not covered at all in any of the questionnaires. In particular, questions about several aspects of patient engagement are missing from all instruments.
- Questions on technical competence are often absent, badly worded or include issues on which patients are not capable of commenting. More work should be done to understand the value of patients' views about technical competence in this type of questionnaire and on the sort of issues they can assess.

In summary, few of the questionnaires are as patient-centred as they ought to be. As the relationship between doctors and patients changes, and patients become more questioning, demanding and involved in their own self-care it is time for a fresh look at instruments designed to obtain patients' feedback on their doctors. We therefore recommend that:

- Questionnaires more attuned to the patient-engagement agendas of today are developed and include a fuller range of questions
- Further exploratory research is carried out to investigate the various elements of physicians' technical competence, patients' capacity to comment on it, and how such findings should be interpreted

- Further consideration is given to the development of questionnaires targeted to specific types of condition or specialty as well as those designed to be administered across a broad range of settings.
- More work is done to determine the best way to administer patient feedback surveys in clinical settings.
- Instructions for implementation are examined further, as it is impossible to tell from most of the published studies what instructions are given[xviii].

High quality patient feedback is important, and those in the business of assessing doctors need to know what makes a good questionnaire. It is a job for experts and takes time and money. There are very many weak questionnaires in existence. Those we have reviewed here have had considerable resources expended on them and so are likely to be among the best available. Even so, several fall short of the ideal, a salutary lesson for those organisations thinking of designing their own.

Like doctors, questionnaires should be fit for their job. Today they are needed more than ever as doctors' conduct, competence and performance become the subject of ever greater scrutiny. We recommend that well-developed questionnaires are routinely used as part of ongoing licensure or certification processes (as the PAR is in Alberta). As the Chief Medical Officer for England has recently said in his report on the regulation of doctors (Chief Medical Officer 2006), because there will always be some poorly performing doctors it is vital to "recognise the problems early … and deal with them effectively by rigorous, fair assessment …" Patients are acknowledged as part of that process, better placed than anyone else to assess some aspects of the doctor's role, and thus to contribute to clinical governance, assessment of trainees, appraisal and revalidation of doctors. They should be given the opportunity to make the assessment with instruments worthy of that task. Only then will patients' views really count.

8 References

ACGME (2005) *Accreditation Council for Graduate Medical Education Outcome Project*.

Arborelius, E. & Bremberg, S. (1992) 'What does a human relationship with the doctor mean?'. *Scandinavian Journal of Primary Health Care*, 10: 163-169.

Avis, M., Bond, M. and Arthur, A. (1995) 'Satisfying solutions? A review of some unresolved issues in the measurement of patient satisfaction'. *Journal of Advanced Nursing*, 22 (2): 316-322.

Baker, R. (1990) 'Development of a questionnaire to assess patients' satisfaction with consultations in general practice'. *British Journal of General Practice*, 40: 487-490.

Baker, R. & Whitfield, M. (1992a) 'Measuring patient satisfaction: a test of construct validity'. *Quality and Safety in Health Care*, 1 (2): 104-109.

Baker, R. & Whitfield, M. (1992b) 'Measuring patient satisfaction: a test of construct validity'. *Quality and Safety in Health Care*, 1 (2): 104-109.

Baker, R. & Whitfield, M. (1992c) 'Measuring patient satisfaction: a test of construct validity'. *Quality and Safety in Health Care*, 1 (2): 104-109.

Beihn, J. & Molineux, J. (1979) 'Patient evaluation of phaysician performance'. *Journal of Family Practice*, 8: 565-569.

BMA Board of Medical Education (2004) *Communication skills education for doctors: an update*. London: British Medical Association.

Bower, P., Mead, N. and Roland, M. (2002a) 'What dimensions underlie patient responses to the General Practice Assessment Survey? A factor analytic study'. *Family Practice*, 19 (5): 489-495.

Bower, P., Mead, N. and Roland, M. (2002b) 'What dimensions underlie patient responses to the General Practice Assessment Survey? A factor analytic study'. *Family Practice*, 19 (5): 489-495.

Bower, P., Mead, N. and Roland, M. (2002c) 'What dimensions underlie patient responses to the General Practice Assessment Survey? A factor analytic study'. *Family Practice*, 19 (5): 489-495.

Bower, P. & Roland, M. (2003) 'Bias in patient assessments of general practice: general practice assessment survey scores in surgery and postal responders'. *British Journal of General Practice*, 53: 126-128.

Bowling, A. (2002) *Research Methods in Health*. Buckingham: Open University Press.

Brownlea, A., Buckley, B., Field, D., Francis, W., Greco, M. and McGovern, J. (1999) *Doctors Advancing Interpersonal Skills (DAIS): A workbook/video educational program.* Australia: CFEP.

Buetow, S. (1995) 'What do general practitioners want from general practice and are they receiving it? A framework'. *Social Science & Medicine*, 40: 213-221.

CFEP (2000) *A pilot study of patient views on the communication skills of general practitioners*. Melbourne, Australia: RACGP Outcomes Evaluation Unit.

Cleary, P.D. (1999) 'The increasing importance of patient surveys'. *Quality and Safety in Health Care*, 8 (4): 212.

Conrad, F. & Couper, M. (2004) *Usability, comparability and data quality across modes and technologies in census data collection.* Gunnison Consulting Group, Inc.

Corrado, M., Carluccio, A., Wilkins, C. and Norton, A. (2005) *Attitudes to Medical Regulation and Revalidation, Survey carried out for the Department of Health*. London: Mori.

Coulter, A. (2006a) *Trends in patients' experience of the NHS*. Oxford: Picker Institute Europe.

Coulter, A. (2006b) 'Can patients assess the quality of health care?'. *BMJ*, 333 (7557): 1-2.

Coulter, A. (2005) 'What do patients and the public want from primary care?'. *BMJ*, 331 (7526): 1199-1201.

Coulter, A. & Ellins, J. (2006) *The quality enhancing interventions project: patient-focused interventions*. London: The Health Foundation.

Coulter, A. & Magee, H. (2003) *The European Patient of the Future*. Maidenhead: Open University Press.

Crossley, J. & Davies, H. (2005) 'Doctors' consultations with children and their parents: a model of competencies, outcomes and confounding influences'. *Medical education*, 39: 807-819.

Crossley, J., Eiser, C. and Davies, H. (2005) 'Children and their parents assessing the doctor-patient interaction: a rating system for doctors' communication skills.'. *Medical education*, 39: 820-828.

Crossley J, Humphris G and Jolly B (2002) 'Assessing health professionals'. *Medical Care*, 36: 800-804.

Davies, H. & Howells, R. (2004) 'How to assess your specialist registrar'. *Archives of Disease in Childhood*, 89: 1089-1093.

Davies, A. & Ware, J. (1988) 'Involving consumers in quality of care assessment'. *Health Affairs*, 7: 33-48.

Delbanco, T. (1992) 'Enriching the doctor-patient relationship by inviting the patient's perspective'. *Annals of Internal Medicine*, 116: 414-418.

Dillman, D. (2002) *Mail and Internet surveys: the tailored design method*. New York: Wiley.

Epstein, R. & Hundert, E. (2002) 'Defining and assessing professional competence'. *Journal of the American Medical Association*, 287: 226-235.

Falvo, D. & Smith, J. (1983) 'Assessing residents' behavioural sciences skills: patients' views of physician-patient interaction'. *Journal of Family Practice*, 17: 479-483.

Frank, J.E. (2005) *The CanMEDs 2005 physician competency framework. Better standards. Better physicians. Better care.* Ottowa: The Royal College of Physicians and Surgeons of Canada.

GMC (2001) *Good Medical Practice*. London: General Medical Council.

Goldman, R. (1994) 'The reliability of peer assessments: a meta-analysis.'. *Evaluations and the health professions*, 17: 3-21.

Greco, M., Spike, N., Powell, R. and Brownlea, A. (2002) 'Assessing communication skills of GP registrars: a comparison of patient and GP examiner ratings'. *Medical education*, 36: 336-376.

Greco, M. & Pocklington, S. (2001) 'Incorporating patient feedback into vocational training: an interpersonal skills development exercise for GP Trainers and Registrars'. *Education for General Practice*, 12 (3): 258-291.

Greco, M., Brownlea, A. and McGovern, J. (2001a) 'Impact of patient feedback on the interpersonal skills of GP Registrars: results of a longitudinal study'. *Medical education*, 36: 336-376.

Greco, M., Sweeny, K., Broomhall, J. and Beasley, P. (2001b) 'Patient assessment of interpersonal skills@ a clinical governance activity for hospital doctors and nurses'. *Journal of Clinical Excellence*, 3 (3): 117-124.

Greco, M., Brownlea, A., McGovern, J. and Cavanagh, M. (2000) 'Consumers as educators: implementation of patient feedback in general practice training'. *Health Communication*, 12 (2): 173-193.

Greco, M., Cavanagh, M., Brownlea, A. and McGovern, J. (1999) 'Validation studies of the Doctors' Interpersonal Skills Questionnaire'. *Education for General Practice*, 10: 256-264.

Greco, M., Francis, W., Buckley, B., Brownlea, A. and McGovern, J. (1998) 'Real-patient evaluation of communication skills teaching for GP Registrars'. *Family Practice*, 15 (1): 51-57.

Greco, M., Brownlea, A. and McGovern, J. (1995) 'Utilising patient feedback in the RACGP Training Program: An exploratory study.'. *Australian Family Physician*, 24 (6): 1077-1081.

Greenhalgh, T. & Eversley, J. (1999) *Quality in General Practice: Towards a Holistic Approach*. London: King's Fund.

Hall, W., Violato, C., Lewkonia, R., Lockyer, J., Fidler, H., Toews, J., Jennett, P., Donoff, M. and Moores, D. (1999) 'Assessment of physician performance in Alberta: the Physician Achievement Review'. *Canadian Medical Association Journal*, 161: 52-57.

Hargraves, J., Hays, R. and Cleary, P. (2003) 'Psychometric properties of the Consumer Assessment of Health Plans Study (CAHPS) 2.0 Adult Core Survey'. *Health Services Research*, 38 (6): 1509-1527.

Hasman, A., Coulter, A. and Askham, J. (2006) *Education for Partnership: developments in medical education*. Oxford: The Picker Institute.

Hays, R., Chong, K., Brown, J., Spritzer, K. and Horne, K. (2003) 'Patient Reports and Ratings of Individual Physicians: An Evaluation of the DoctorGuide and Consumer Assessment of Health Plans Study Provider-Level Surveys'. *American Journal of Medical Quality*, 18 (5): 190-196.

Hearnshaw, H., Baker, R., Cooper, A., Eccles, M. and Soper, J. (1996) 'The costs and benefits of asking patients their opinions about general practice.'. *Family Practice*, 13: 52-58.

Howie, J., Heaney, D., Maxwell, M., Walker, J. and Freeman, G. (2000) 'Developling a 'consultation quality index' (CQI) for use in gereral practice'. *Family Practice*, 17 (6): 455-461.

Howie, J., Heaney DJ, Maxwell M and Walker JJ (1998) 'A comparison of a Patient Enablement Instrument (PEI) against two established satisfaction scales as an outcome measure of primary care consultations'. *Family Practice*, 15 (2): 165-171.

Hutchinson, A., Williams, M., Meadows, K., Barbour, R.S. and Jones, R. (1999) 'Perceptions of good medical practice in the NHS: a survey of senior health professionals'. *Quality and Safety in Health Care*, 8 (4): 213-218.

Institute of Medicine (1994) *Defining Primary Care. An interim report*. Washington, DC: National Academy Press.

Jung, H., Wensing, M. and Grol, R. (1997) 'What makes a good general practitioner: do patients and doctors have different views?'. *British Journal of General Practice*, 47: 805-809.

Kaplan, S. & Ware, J. (1989) 'The patient's role in healht care and quality assessment'. In Goldfield, N. & Nash, D. (eds), *Providing Quality Care: The challenge to clinicians*. Philadelphia: Wiley and Sons.

Klessig, J., Robbins, A., Wienland, D. and Rubenstein, L. (1989) 'Evaluating humanistic attributes of internal medicine residents'. *Journal of General Internal Medicine*, 4: 514-521.

Kurtz, S. et al (1998) *Teaching and learning communication skills in medicine*. Oxford: Radcliffe Medical Press.

Levine, A. (2002) 'Medical professionalism in the new millenium: A physician charter'. *Annals of Internal Medicine*, 136: 243-226.

Link, M. & Mokdad, A. (2005) 'Alternative modes for health surveillance surveys: An experiment with web, mail and telephone'. *Epidemiology*, 40 (6): 2120-2139.

Lipner, R., Blank, L., Leas, B. and Fortna, G. (2002) 'The value of patient and peer ratings in recertification'. *Academic Medicine*, 77 (10 Supplement): S64-S66.

Marshall, G. & Hays, R. (1994) *The Patient Satisfaction Questionnaire Short-Form (PSQ-18)*. Santa Monica, CA: RAND.

Marshall, G., Hays, R., Sherbourne, C. and Wells, K. (1993) 'The structure of patient satisfaction with outpatient medical care'. *Psychological Assessment*, 5: 477-483.

Maudsley, R., Wilson, D., Neufield, V., Hennen, B., DeVillaer, M. and Wakefield, J. (2000) 'Educating future physicians for Ontario: phase II'. *Academic Medicine*, 75: 113-126.

McKinstry, B., Walker, J., Blaney, D., Heaney, D. and Begg, D. (2004) 'Do patients and expert doctors agree on the assessment of consultation skills?'. *Family Practice*, 21 (1): 75-80.

McLoed, P. & Tamblyn, R. (1994) 'Faculty ratings of resident humanism predict patient satisfaction ratings in abmulatory medical clinics'. *Journal of General Internal Medicine*, 9: 321-326.

McWhinney, I. (1989) *A textbook of family medicine*. New York: Oxford University Press.

Mercer, S., McConnachie, A., Maxwell, M., Heaney, D. and Watt, G. (2005) 'Relevance and practical use of the Consultation and Relational Empathy (CARE) measure in general practice'. *Family Practice*, 22: 328-334.

Mercer, S., Maxwell, M., Heaney, D. and Watt, G. (2004) 'The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure'. *Family Practice*, 21 (6): 699-705.

Mercer, S., Reilly, D. and Watt, G. (2002) 'The importance of empathy in the enablement of patients attending the Glasgow Homeopathic Hospital.'. *British Journal of General Practice*, 52: 901-905.

Merkel, W. (1984) 'Physician perception of patient satidfaction: do doctors know which patients are satisfied?'. *Medical Care*, 22: 453-459.

Merterko, M., Kosinski, M. and Davies, A. (1994) *Physician-specific patient satisfaction profiles: Report of a pilot study at Tufts Associated Health Plans*. New England Medical Center.

National Primary Care Research and Development Centre (NPCRDC) (2004) *Manual General Practice Assessment Questionnaire (GPAQ) Version 1.5*. Manchester: University of Manchester.

Nelson, E., Gentry, M., Mook, K., Spritzer, K., Higgins, J. and Hays, R.D. (2006) 'How many patients are needed to provide reliable evaluations of individual clinicians?'. *Medical Care*, 42 (3): 259-266.

Ramsay, J., Campbell, JL., Schroter, S., Green, J. and Roland, M. (2000a) 'The General Practice Assessment Survey (GPAS): tests of data quality and measurement properties'. *Family Practice*, 17 (5): 372-379.

Ramsay, J., Campbell, JL., Schroter, S., Green, J. and Roland, M. (2000b) 'The General Practice Assessment Survey (GPAS): tests of data quality and measurement properties'. *Family Practice*, 17 (5): 372-379.

Ramsay, J., Campbell, JL., Schroter, S., Green, J. and Roland, M. (2000c) 'The General Practice Assessment Survey (GPAS): tests of data quality and measurement properties'. *Family Practice*, 17 (5): 372-379.

Ramsay, J., Campbell, JL., Schroter, S., Green, J. and Roland, M. (2000d) 'The General Practice Assessment Survey (GPAS): tests of data quality and measurement properties'. *Family Practice*, 17 (5): 372-379.

Rao, M., Clarke, A., Sanderson, C. and Hammersley, R. (2006) 'Patients' own assessments of quality of primary care compared with objective records based measures of technical quality of care: cross sectional study'. *BMJ*, 333 (7557): 19.

Rees-Lewis, J. (1994) 'Patients' views on quality care in general practice: literature review'. *Social Science & Medicine*, 39: 655-671.

Rodriguez, H., von Glahn, T., Rogers, W., Chang, H., Fanjiang, G. and Safran, D. (2006) 'Evaluating Patients' Experiences with Individual Physicians: A Randomized Trial of Mail, Internet, and Interactive Voice Response Telephone Administration of Surveys.'. *Medical Care*, 44 (2): 167-174.

Roter, D., Stewart, M., Putnam, S., Lipkin, M., Stiles, W. and Inui, T. (1997) 'Communication patterns of primary care physicians'. *Journal of the American Medical Association*, 277: 350-356.

Safran, D., Kosinski, M., Tarlov, A.R., Rogers, W.H., Taira, D.A., Lieberman, N. and Ware, J.E. (1998) 'The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance'. *Medical Care*, 36 (5): 728-739.

Smith, C. & Armstrong, D. (1989) 'Comparison of criteria derived by government and patients for evaluating general practitioner services'. *British Medical Journal*, 299: 494-496.

Southgate, L. (1994) 'Freedom and discipline: clinical practice and the assessment of clinical competence'. *British Journal of General Practice*, 44: 87-92.

Stewart, M. (1995) 'Effective physician-patient communication and health outcomes: a review'. *Canadian Medical Association Journal*, 152: 1423-1433.

Suchman, A., Roter, D., Green, M. and Lipkin, M. (1993) 'Physician satisfaction with primary care office visits. Collaborative Study Group of the American Academy on Physician and Patient.'. *Medical Care*, 31: 1083-1092.

Swanson, D. et al (1990) 'Precision of patients' ratings of residents' humanistic qualities: How many items and patients are enough?'. In Bender, W. et al (eds), *Teaching and Assessing Clinical Competence*. Groningen, the Netherlands.

van Dulmen, S. (2003) 'Patient-centredness'. *Patient Education and Counseling*, 51 (3): 195-196.

Violato, C., Lockyer, J. and Fidler, H. (2003) 'Multisource feedback: a method of assessing surgical practice'. *British Medical Journal*, 326: 546-548.

Violato, C., Marini, A., Toews, J., Lockyer, J. and Fidler, H. (1997) 'Using peers, self, patients and co-workers to assess physician performance'. *Academic Medicine*, 72: 582-584.

Ware, J., Snyder, N., Wright, W. and Davies, A. (1983) 'Defining and measuring patient satisfaction with medical care'. *Evaluation and Program Planning*, 6: 247-263.

Ware, J., Davies-Avery, A. and Stewart, A. (1978) 'The measurement and meaning of patient satisfaction'. *Health and Medical Care Services Review*, 1: 1-15.

Ware, JE, Effects of differences in quality of care on patient satisfaction. Proceedings of the 17th Annual Conference on Research in Medical Education. CONFERENCE PROCEEDING

Ware, J., Snyder, N. and Wright, W. (1976a) *Development and validation of scales to measure patient satisfaction with medical care services Vol1, Part A: Review of literature, overview of methods and results regarding construction of scales.* Springfield, VA: National Technical Information Service.

Ware, J., Snyder, N. and Wright, W. (1976b) *Development and validation of scales to measure patient satisfaction with medical care services Vol1, Part B: Results regarding scales constructed from the Patient Satisfaction Questionnaire and measure of other health care perceptions.* Springfield, VA: National Technical Information Service.

Weaver, M., Ow, C., Walker, D. and Degenhardt, E. (1993) 'A questionnaire for patients' evaluations of their physicians' humanistic behaviors'. *Journal of General Internal Medicine*, 8 (3): 135-139.

Webster, G. (1989) *Final report on the Patient Satisfaction Questionnaire Project*. Philadelphia: American Board of Internal Medicine.

Wensing M, Jung HP, Mainz J, Olesen F and Grol R (1998) 'A systematic review of the literature on patient priorites for general practice care. Part 1: Describtion of the research domain'. *Soc.Sci.Med.*, 47 (10): 1573-1588.

Wilson, R., Harrison, B., Gibberd, R. and Hamilton, J. (1999) 'An analysis of adverse events from the quality in Australian health care study'. *Medical Journal of Australia*, 170: 411-415.

Wolliscroft, J. & Howell, J. (1994) 'Resident-patient interactions: the humanistic qualities of internal medicine residents assessed by patients, attending physicians, program supervisors and nurses'. *Academic Medicine*, 69: 216-214.

Zaslavsky, A.M. & Cleary, P.D. (2002) 'Dimension of plan performance for sick and healthy members on the Consumer Assessments of Health Plans Study 2.0 survey'. *Medical Care*, 40 (10): 951-964.

Zaslavsky, A.M., Beaulieu, N.D., Landon, B.E. and Cleary, P.D. (2000) 'Dimensions of consumer-assessed quality of Medicare managed-care health plans'. *Medical Care*, 38 (2): 162-174.

9 Appendices

Appendix 1: Details of topics included in the selected questionnaires

General Practice Assessment Questionnaire (GPAQ)

The GPAQ covers four pages, including 19 (post-consultation version) or 20 (postal version) items with many sub-items, including three free text questions.

It includes many items about access and organisation, as well as items about interpersonal skills, communication and engagement. Both post-consultation and postal versions cover the following topics:

Receptionists; opening hours; waiting times for an appointment with a particular doctor or any doctor; waiting times in waiting room; getting through and speaking to a doctor on the phone; continuity; asking about symptoms/how patient is feeling; listening; put at ease; involvement in decisions; explanation; time spent; patience; caring and concern.

The consultation version of the GPAQ includes three items on 'enablement' adapted from the Patient Enablement Instrument (see section 5.10), asking *"after seeing the doctor today, do you feel…*

Able to understand and cope with problem(s) or illness

Able to cope with your problem(s) or illness

Able to keep yourself healthy

with response options: *"much more than before the visit," "a little more than before the visit," "the same or less than before the visit"* and *"does not apply."* The authors report that these items were less well understood in a postal context (National Primary Care Research and Development Centre (NPCRDC) 2004) so they were included only in the post-consultation version.

Five items in the consultation version of GPAQ and nine from the postal version cover issues unrelated to the individual doctor's performance, measuring either organisational aspects of the practice or receptionists/nurses' performance. Unlike the first version of GPAS, it includes no indicators of technical care. Questions about technical competence (eg GPs' diagnostic or management skills) were removed from version two of the GPAS since qualitative interviews with patients indicated that the validity of patients' judgement of doctors' technical competence is low (Bower et al 2002c).

The sociodemographics section of GPAS includes an item about limiting long-standing illness, worded in the same way as 2001 census, which is a strong predictor of a high consultation rate. It also includes a condensed version of a census question about employment status. These allow comparisons with local and wider populations.

Interpersonal skills		Communication of information		Patient engagement and enablement		Global satisfaction	
Instilling confidence/ trust/ being open	✓	Quality of explanations	✓	Help to understand illness, health and treatment	×	Overall satisfaction	✓
Holistic, empathic	×	Eliciting info/ concerns from patient	✓	Involvement in decisions	✓	Would recommend doctor to friend or family	×
Caring, polite, good with people	✓	Thoroughness of discussion	×	Provision of preventive advice	✓	Would choose this doctor in future	×
Respect, privacy, dignity	×	Listening	×	Information about medication, etc	✓		
				Access to other sources of info/support	×		
				Risk communication	×		
				Promoting health literacy	×		

Doctors' Interpersonal Skills Questionnaire (DISQ)

DISQ contains 12 items on one page, and covers the following topics: greeting; listening; explanation; eliciting concerns; reassurance; time; consideration of personal situation; respect; privacy and dignity; recommendation to friends.

DISQ asks patients to rate the doctor's performance on a five-point scale from *"poor"* to *"excellent"* on these areas:

"The warmth of this doctor's greeting to me was.."

"The respect shown to me by this doctor was…"

"The doctor's concern for me as a person in this visit was…"

DISQ includes two items which assess the doctor's capacity to ask about and listen to the patient:

"On this visit I would rate the doctor's ability to really listen to me as…"

"The opportunity the doctor gave me to express my concerns or fears was…"

DISQ asks patients to rate whether the doctor takes a holistic approach to the consultation:

"This doctor's consideration of my personal situation in deciding a treatment or advising me was…"

DISQ includes a well-constructed single global satisfaction item, answered on a five-point likert scale from *"poor"* to *"excellent."*

"The recommendation I would give to my friends about this doctor would be.."

Interpersonal skills		Communication of information		Patient engagement and enablement		Global satisfaction	
Instilling confidence/ trust/ being open	✓	Quality of explanations	✓	Help to understand illness, health and treatment	×	Overall satisfaction	✓
Holistic, empathic	✓	Eliciting info/ concerns from patient	✓	Involvement in decisions	×	Would recommend doctor to friend or family	✓
Caring, polite, good with people	✓	Thoroughness of discussion	✓	Provision of preventive advice	×	Would choose this doctor in future	×
Respect, privacy, dignity	✓	Listening	✓	Information about medication, etc	×		
				Access to other sources of info/support	×		
				Risk communication	×		
				Promoting health literacy	×		

Consultation Satisfaction Questionnaire (CSQ)

CSQ contains 18 items which are statements about four dimensions of patient satisfaction with the most recent consultation: general satisfaction; professional care; depth of relationship; perceived time spent with doctor.

It includes one measure of patient engagement, by asking patients their level of agreement with:

"I understand my illness better after seeing this doctor"

It also reflects the value to patients of a holistic approach to their care:

"This doctor was interested in me as a person and not just my illness."

Three of the CSQ's 18 items asks directly for an evaluation of quality of consultation or overall satisfaction. This seems an unwise use of space in a questionnaire designed to gather feedback for formative purposes.

"I am totally satisfied with my visit to/from this doctor"

"I am not completely satisfied with my visit to/from this doctor"

"Some things about the consultation with the doctor could have been better"

Interpersonal skills		Communication of information		Patient engagement and enablement		Global satisfaction	
Instilling confidence/ trust/ being open	✓	Quality of explanations	✓	Help to understand illness, health and treatment	✓	Overall satisfaction	✓
Holistic, empathic	✓	Eliciting info/ concerns from patient	×	Involvement in decisions	×	Would recommend doctor to friend or family	×
Caring, polite, good with people	✓	Thoroughness of discussion	✓	Provision of preventive advice	×	Would choose this doctor in future	×
Respect, privacy, dignity	×	Listening	✓	Provision of information about medication, etc	×		
				Access to other sources of info/support	×		
				Risk communication	×		
				Promoting health literacy	×		

CARE Measure

The CARE instrument consists of ten items on one page, which ask patients to rate doctors on a five-point likert scale from *"poor"* to *"excellent,"* asking for each *"How was the doctor at.."*

Making you feel at ease; Letting you tell your 'story'; Really listening; being interested in you as a whole person; fully understanding your concerns; showing care and compassion; being positive; explaining things clearly; helping you take control; making a plan of action with you.

The last two items are related to enablement.

It also asks how important the items were in the context of this consultation. This is a good question, but the value of asking it once in relation to all the questions is dubious. It may be more useful to ask it once for each item, although this would have space implications.

Interpersonal skills		**Communication of information**		**Patient engagement and enablement**		**Global satisfaction**	
Instilling confidence/ trust/ being open	×	Quality of explanations	✓	Help to understand illness, health and treatment	×	Overall satisfaction	×
Holistic, empathic	✓	Eliciting info/ concerns from patient	✓	Involvement in decisions	✓	Would recommend doctor to friend or family	×
Caring, polite, good with people	✓	Thoroughness of discussion	×	Provision of preventive advice	✓	Would choose this doctor in future	×
Respect, privacy, dignity	×	Listening	✓	Provision of information about medication, etc	✓		
				Access to other sources of info/support	×		
				Risk communication	×		
				Promoting health literacy	×		

SHEFFPAT

The questionnaire covers two pages, and is made up of 13 items plus demographics and a space for other comments. It asks about consultation quality: opportunity to shape discussion; quality of advice; patient's understanding of condition and treatment; confidence in self-care; interpersonal skills; the doctor's interest in patient's point of view; quality of explanation; respect; confidentiality; overall satisfaction; demographics.

The items are based on *Good Medical Practice* section on 'communication with patients' so is likely to be suitable for revalidation and in-training assesment (Crossley et al, 2005)

Interpersonal skills		Communication of information		Patient engagement and enablement		Global satisfaction	
Instilling confidence/ trust/ being open	×	Quality of explanations	✓	Help to understand illness, health and treatment	×	Overall satisfaction	✓
Holistic, empathic	×	Eliciting info/ concerns from patient	×	Involvement in decisions	✓	Would recommend doctor to friend or family	×
Caring, polite, good with people	✓	Thoroughness of discussion	×	Provision of preventive advice	×	Would choose this doctor in future	×
Respect, privacy, dignity	✓	Listening	✓	Provision of information about medication, etc	✓		
				Access to other sources of info/support	×		
				Risk communication	×		
				Promoting health literacy	×		

ABIM Questionnaire

The questionnaire covers two pages, contains 18 items and takes eight minutes to complete by telephone. It covers the following issues: Truthfulness; manner; not being patronising; listening; showing interest; information during physical examination; involvement in decision-making; encouraging and answering questions; giving information about problems; clear explanations; kind of doctor; length of time under doctor's care; number of visits; whether the patient would recommend this doctor to others. It also asks for some demographic information about the patient: age; sex; health status.

Items within the ABIM questionnaire address interpersonal skills. For example, *"How is this doctor at…*

Greeting you warmly, calling you by the name you prefer, being friendly, never crabby or rude.

Treating you like you're on the same level, never "talking down" to you or treating you like a child.

Many items in the ABIM questionnaire have to do with giving information, such as *how is this doctor at*

Telling you everything, being truthful, upfront and frank; not keeping things from you that you should know.

Warning you during the physical exam about what he/she is going to do and why; telling you what he/she finds.

Explaining what you need to know about your problems and treatment; explaining any technical medical terms in plain language."

ABIM questionnaire recognises the importance of a doctor's capacity to elicit, listen to and understand information from patients as well as to give information to patients. It is very thorough in this area asking, for example, *How is this doctor at…*

Letting you tell your story; listening carefully; asking thoughtful questions; not interrupting you while you're talking.

Interpersonal skills		Communication of information		Patient engagement and enablement		Global satisfaction	
Instilling confidence/ trust/ being open	✓	Quality of explanations	✓	Help to understand illness, health and treatment	✓	Overall satisfaction	×
Holistic, empathic	×	Eliciting info/ concerns from patient	✓	Involvement in decisions	✓	Would recommend doctor to friend or family	✓
Caring, polite, good with people	✓	Thoroughness of discussion	✓	Provision of preventive advice	×	Would choose this doctor in future	×
Respect, privacy, dignity	×	Listening	×	Provision of information about medication, etc	×		
				Access to other sources of info/support	×		
				Risk communication	×		
				Promoting health literacy	×		

PAR Programme Questionnaires

The precise content of the PAR programme questionnaires varies depending on the specialism with which it is designed to be used and the respondent group (patient, co-worker and peer) but generally the questionnaires cover five attributes of the physician's performance: clinical knowledge and skills; communication skills; psychosocial management; office management; collegiality (http://www.par-program.org/PAR-History2.htm accessed 26.4.2006

The length of the patient questionnaires varies between specialisms. For example, questionnaires for primary care physicians, medical specialists and surgeons are two pages, with 39 or 40 items. Questionnaires for use in anaesthesia or episodic care are one page, with 11 or 16 items respectively.

The attributes measured by the General Practitioner patient questionnaire are: Patient Interaction (the physician listens, answers questions and demonstrates interest, empathy and respect for the patient during an examination period. Patients indicate whether they would return to or refer a friend to the physician); phone communication (the availability of a doctor by phone after hours for urgent medical problems); personal communication (the physician adequately explains illness/injury, preventative measures, treatment options, and medication regimen and side effects); information for patients (the physician provides proper information regarding medical problems, return appointments, reporting of test results, referrals to specialists, tracking of prescription and non-prescription medication and patient education); office staff (the staff is pleasant, helpful, capable, professional and able to maintain confidentiality); physical office (the office is accessible, clean, private, and appropriately sized); appointments (appointments can be made quickly and wait time for scheduled appointments is not excessive).

The questionnaires can all be accessed at http://www.par-program.org/PAR-Inst.htm (last accessed 26.4.2006) and the groups of attributes each measures can be found here: http://www.par-program.org/PAR-Attrib.htm (last accessed 26.4. 2006)

Interpersonal skills		Communication of information		Patient engagement and enablement		Global satisfaction	
Instilling confidence/ trust/ being open	×	Quality of explanations	✓	Help to understand illness, health and treatment	✓	Overall satisfaction	×
Holistic, empathic	×	Eliciting info/ concerns from patient	✓	Involvement in decisions	✓	Would recommend doctor to friend or family	✓
Caring, polite, good with people	✓	Thoroughness of discussion	×	Provision of preventive advice	✓	Would choose this doctor in future	✓
Respect, privacy, dignity	✓	Listening	×	Provision of information about medication, etc	✓ ✓ ✓		
				Access to other sources of info/support	×		
				Risk communication	×		
				Promoting health literacy	×		

Patient Satisfaction Questionnaire (PSQ-18) – short form

The two-page PSQ-18 retains many characteristics of the PSQ III. Its 18 items measure each of the seven dimensions of satisfaction with medical care measured by the PSQ III (Marshall and Hays, RD 1994): general satisfaction; technical quality; interpersonal manner; communication; financial aspects; time spent with doctor; accessibility and convenience.

Interpersonal skills		Communication of information		Patient engagement and enablement		Global satisfaction	
Instilling confidence/ trust/ being open	✓	Quality of explanations	✓	Help to understand illness, health and treatment	×	Overall satisfaction	✓
Holistic, empathic	×	Eliciting info/ concerns from patient	×	Involvement in decisions	×	Would recommend doctor to friend or family	×
Caring, polite, good with people	✓	Thoroughness of discussion	×	Provision of preventive advice	×	Would choose this doctor in future	×
Respect, privacy, dignity	×	Listening	✓	Provision of information about medication, etc	×		
				Access to other sources of info/support	×		
				Risk communication	×		
				Promoting health literacy	×		

CAHPS 2.0

The questionnaire contains 43 items, 19 of which are core items. These include two global ratings of care, and also produce composite scores for five areas of care: getting care quickly; doctors who communicate well; courteous/helpful office staff; getting needed care; Health plan customer service. The remaining questions ask about health plan usage, demographics or are screening questions.

The physician instrument included four items assessing access to care, eight items assessing communication and two preventive care items.

Items which address doctor-patient communication are included in both versions of CAHPS:

In the last 12 months, how often did doctors or other health professionals explain things in a way you could understand?

In the last 12 months, how often did doctors or other health professionals listen carefully to you?

In the last 12 months, how often did doctors or other health professionals show respect for what you had to say?

Interpersonal skills		**Communication of information**		**Patient engagement and enablement**		**Global satisfaction**	
Instilling confidence/ trust/ being open	×	Quality of explanations	✓	Help to understand illness, health and treatment	×	Overall satisfaction	✓
Holistic, empathic	×	Eliciting info/ concerns from patient	×	Involvement in decisions	✓	Would recommend doctor to friend or family	✓
Caring, polite, good with people	×	Thoroughness of discussion	×	Provision of preventive advice	✓	Would choose this doctor in future	×
Respect, privacy, dignity	✓	Listening	✓	Provision of information about medication, etc	✓		
				Access to other sources of info/support	×		
				Risk communication	×		
				Promoting health literacy	×		

Patient Enablement Instrument (PEI)

The questionnaire includes only six items on one page, each of which asks about some aspect of patient enablement.

The PEI focuses specifically on outcomes related to themes of patient-centredness and empowerment: issues related to, but conceptually distinct from, patient satisfaction (Howie et al 2000, Howie et al 1998). The authors define 'patient-centredness' as an indication by doctors of a commitment to value patients' contribution to deciding what is wrong with them and how their care should be managed. Empowerment is taken to mean that patients are helped to understand the nature of their problems and enabled to manage their own illness. Response options for the PEI are *much better/Much more, better/more* or *same or less,* and the items ask, *As a result of your visit to the doctor today, do you feel you are...*

able to cope with life

able to understand your illness

able to cope with your illness

able to keep yourself healthy

confident about your health

able to help yourself

Interpersonal skills		**Communication of information**		**Patient engagement and enablement**		**Global satisfaction**	
Instilling confidence/ trust/ being open	×	Quality of explanations	×	Help to understand illness, health and treatment	×	Overall satisfaction	×
Holistic, empathic	×	Eliciting info/ concerns from patient	×	Involvement in decisions	×	Would recommend doctor to friend or family	×
Caring, polite, good with people	×	Thoroughness of discussion	×	Provision of preventive advice	✓	Would choose this doctor in future	×
Respect, privacy, dignity	×	Listening	×	Provision of information about medication, etc	✓		
				Access to other sources of info/support	×		
				Risk communication	×		
				Promoting health literacy	×		

Appendix 2: Details of evaluations of measurement characteristics of each instrument

General Practice Assessment Questionnaire (GPAQ)

Measurement characteristics

The psychometric properties of GPAQ itself have not been tested. However, GPAQ differs only slightly from GPAS, the previous version of the questionnaire GPAS, which was tested extensively, and the authors claim the properties of the two questionnaires are likely to be similar. Further, the GPAQ has been cognitively tested.

Reliability

A number of aspects of development of GPAS were clearly explained and undertaken thoroughly. The evaluation was based, for the most part, on large samples, and appropriate statistics were used. The internal consistency of multiple item scales was rigorously tested, and Cronbach's alpha reported (alpha>0.70 for each scale other than trust, where alpha=0.69).

However, the findings are not always clearly reported. Data which assessed the homogeneity of the measure and completeness of data were supplied and the testing was reasonably rigorous.

The physician-level reliability for the assessment of individual practitioners was not explored

Validity

The authors claim that the construct validity of GPAS is supported by the three types of within-scale analysis (Ramsay et al, 2000), but no evaluation of the criterion or construct validity against external criteria is reported.

Measurement characteristics			
Internal consistency	✓	Criterion validity	×
Factor structure	✓	Construct validity	×
Physician-level reliability	×		

Doctors' Interpersonal Skills Questionnaire (DISQ)

Reliability

Internal reliability was tested and found to be high (Cronbach's alpha =0.96)

Principal components analysis was carried out and demonstrated that DISQ is a single-scale instrument (Eigenvalue= 8.44).

The physician-level reliability for the assessment of individual practitioners was not explored.

Validity

Extensive tests of validity were carried out.

Content validity was achieved by involving focus groups of patients and GPs in the development of the questionnaire, as well as a review of literature.

Construct (discriminant) validity: the DISQ has been shown to detect changes in doctors' communication skills before and after training. For example, a five-hour teaching workshop was shown to increase DISQ scores by 3-5% (Greco et al 1998) and in a study by the Royal Australian College of General Practitioners (RACGP) Outcomes Evaluation Unit, DISQ could detect differences between GPs who were and were not vocationally trained (CFEP 2000).

Criterion validity was established by a significant correlation ($r=0.77$, $p<0.0001$) between DISQ scores and the Falvo-Smith Interaction Scale (Falvo & Smith 1983).

Two further studies (Greco et al 2002, Greco et al 1999) tested concurrent validity by demonstrating significant correlations between DISQ scores and ratings of expert GPs and GP examiners. These findings are interesting, given that other studies have demonstrated no correlation between patient and professional ratings of communication skills (Klessig et al 1989, Merkel 1984, Beihn and Molineux, JE 1979)

Measurement characteristics			
Internal consistency	✓	Criterion validity	✓
Factor structure	✓	Construct validity	✓
Physician-level reliability	×		

Consultation Satisfaction Questionnaire (CSQ)

A great deal of work appears to have been undertaken to develop this instrument, but it is not always reported in sufficient detail to be evaluated.

Reliability

Internal reliability for the questionnaire overall was assessed (alpha=0.91). Principal components analysis was carried out and revealed three factors: professional care (alpha=0.87), depth of relationship (alpha=0.83) and perceived time (alpha=0.82) It reported item-total correlations. (Baker, 1990)

The physician-level reliability for the assessment of individual practitioners was not explored.

Validity

The face validity of the instrument is compromised by the ambiguity of some of the items, particularly those measuring 'depth of relationship'. This may reflect the paucity of direct patient involvement in the development of the instrument.

Construct validity and sensitivity were assessed in a study which compared satisfaction, as measured by CSQ, of two sets of patients: a group who had changed their GP practice without changing their home address, with a group who had not (Baker & Whitfield

1992c). The instrument, and the subscales within it, were shown to be sensitive to these changes and, thus, demonstrated construct validity.

Measurement characteristics			
Internal consistency	✓	Criterion validity	×
Factor structure	✓	Construct validity	✓
Physician-level reliability	×		

CARE Measure

Some aspects of the development of this questionnaire appear appropriate, but more information is needed. The content of CARE is carefully constructed on the basis of interviews with patients, and then confirmed or refined in further interviews. However, the statistical analysis of the measurement characteristics is based on very small numbers – much more statistical evidence is required.

Reliability

The internal reliability of the instrument was measured in a series of pilot studies (Cronbach's alpha=0.93).

Scaling assumptions were made without undertaking any factor analysis. It is assumed that the ten items can be summed. Alpha provides some evidence for this belief, but more data are required to be confident.

Generalisability theory was used to estimate the physician-level reliability, which showed a sample of 50 patients is sufficient to obtain a reliable estimate of their mean CARE measure score.

Validity

Face and content validity were developed through a thorough, iterative process of interviews and revisions with a range of patients, who endorsed the clarity and importance of the items in the final instrument. Tested in both high and low deprivation settings, 76% of patients and 78% of doctors rated the items in the instruments as 'very important' to their current consultation (Mercer et al 2005). The high response rate (70% in Mercer et al, 2005) also supports the face validity of the instrument.

Criterion validity of the instruments is assessed by comparing the instrument to two other measures (BLESS and RES) (Mercer et al 2004).

The authors' claims of construct validity (Mercer et al 2004) are based on the views of GPs and expert researchers (not on findings that reflect what would be expected based on research or theory), thus do not demonstrate construct validity.

Measurement characteristics			
Internal consistency	✓	Criterion validity	✓
Factor structure	×	Construct validity	×
Physician-level reliability	✓		

SHEFFPAT

Reliability

Generalisability analysis indicated strong internal reliability. It was also used to estimate physician-level reliability, measuring the reliability of the instrument for discriminating between doctors. It demonstrated that 25 consultations are sufficient for parents' feedback to meet the criteria required for inclusion in a performance assessment programme (reliability of >0.8 using generalisability) (Davies and Howells, R 2004); (Crossley et al 2005)

Factor structure is tested, although the results have not been published.

Validity

Items for SHEFFPAT were derived from an exhaustive literature review and consultation with doctors. Patients did not have input to this process, but did have extensive input to the development process through cognitive interviews.

Construct validity was demonstrated (unpublished data) by the confirmation of three hypotheses which related the accuracy of questionnaire responses to other measures of the same domains.

There is some unpublished evidence of criterion validity which demonstrates a strong correlation between the 'communication with patients' item on the SPRAT multisource assessment instrument and the SHEFFPAT score.

Measurement characteristics			
Internal consistency	✓	Criterion validity	✓
Factor structure	✓	Construct validity	✓
Physician-level reliability	✓		

ABIM Questionnaire

It is interesting and encouraging that a questionnaire for recertification includes patient feedback, but there is little evidence to support the inclusion of the items selected. No evidence is provided for any aspect of validity or reliability.

Measurement characteristics			
Internal consistency	×	Criterion validity	×
Factor structure	×	Construct validity	×
Physician-level reliability	×		

PAR Programme Questionnaires

Reliability

Cronbach's alpha co-efficients were high (over 0.90) for all PAR instruments which provides evidence for internal reliability (Hall et al 1999).

Factor analysis was carried out and retained seven factors which accounted for 70% of the variance. These factors were cohesive, meaningful and provided a multidimensional assessment of their physician.

Physician-level reliability was assessed, and it was found that stability of patient responses ($Ep^2 \geq 0.70$) could be achieved with 23 to 25 patients with 28 patients providing $Ep^2 = 0.84$.

Validity

Violato and Hall (2000) evaluate the concurrent validity of PAR in terms of the extent to which there are correlations between self, patient, peer, consultant and co-worker assessments, but this is problematic as it assumes that other groups' assessments are valid measures of physicians' performance. Concurrent validity should be assessed by comparing the instrument's performance with another validated instrument.

Measurement characteristics			
Internal consistency	✓	Criterion validity	×
Factor structure	✓	Construct validity	×
Physician-level reliability	✓		

Patient Satisfaction Questionnaire (PSQ-18) – short form

Reliability

The measurement characteristics of the PSQ III were extensively demonstrated by (Marshall et al 1993).

The internal consistency of all seven PSQ-18 subscales was assessed, and they ranged from 0.64 to 0.77. Most were above the 0.7 level, although the coefficients for the interpersonal and communication scales fell below 0.70 (Marshall and Hays, RD 1994). The internal reliability of the whole scale is not stated.

Physician-level reliability is not assessed.

Validity

The PSQ-18 subscales correlate significantly with the PSQ III subscales ($r > 0.9$, with one exception), demonstrating criterion validity.

Measurement characteristics			
Internal consistency	✓	Criterion validity	×
Factor structure	✓	Construct validity	×
Physician-level reliability	×		

CAHPS 2.0

Most aspects of development of this questionnaire were rigorously undertaken and clearly explained. Considerable validation work was undertaken.

Reliability

Scaling assumptions were assessed using confirmatory factor analysis, and internal consistency of the five scales was reported. Cronbach's alpha was >0.75 for two of the five subscales, but ranged from 0.58 to 0.62 on the other three scales (Hargraves et al, 2003). These are acceptable levels of internal consistency.

The physician-level reliability is calculated and used to determine the sample size needed. (Hays et al 2003) estimated the number of respondents required to achieve acceptable levels of physician-level reliability for each of three subscales. To achieve a reliability of 0.7 at the physician level for the access to care, communication and preventive care scales, the necessary number of patient responses per physician were 32, 43 and 38 respectively.

Validity

Multivariate associations of composites measures with global ratings are also examined to assess construct validity (Hargraves et al, 2003), but there is no evidence of evaluation against external criteria.

The face validity of the CAHPS 2.0 survey was sought by undertaking thorough field studies and cognitive interviews.

Hays et al (2003) found positive and substantial correlation between CAHPS and DoctorGuide, demonstrating criterion validity of CAHPS.

Measurement characteristics			
Internal consistency	✓	Criterion validity	✓
Factor structure	✓	Construct validity	×
Physician-level reliability	✓		

Patient Enablement Instrument (PEI)

The comparison of the PEI with other measures is good, but other methods are reported somewhat superficially.

Reliability

Internal reliability is assessed rigorously: alpha = 0.93 (Howie et al 1998)

Validity

Criterion validity is rigorously assessed with CSQ and MISS.

Measurement characteristics			
Internal consistency	✓	Criterion validity	✓
Factor structure	×	Construct validity	×
Physician-level reliability	×		

Appendix 3: Summary of strengths and limitations of each questionnaire

	Strengths	Limitations
GPAQ	• Includes three E&E items	• Validity not rigorously demonstrated • No physician-level reliability estimate
DISQ	• Rigorously tested	• Includes no E&E questions • No physician-level reliability estimate
CSQ	• Reliability and construct validity rigorously tested	• Poor wording of items • Only 1 E&E item • No estimate of physician-level reliability
CARE measure	• Includes 3 E&E items • Developed with extensive input from patients • Estimates physician-level reliability	• Construct validity not tested • Factor structure not tested
SHEFFPAT	• Based on 'communication with patients' section of *Good Medical Practice* • Includes 2 E&E items • Estimates physician-level reliability	• Has not yet been validated with adults
ABIMQ	• 2 E&E items • Phone-based administration	• Poor account of development process • Poor wording – multiple item questions • Physician-level reliability not estimated
PAR	• E&E items • Strong development • Estimates physician-level reliability	• Validity testing could be more rigorous
PSQ-18	• Reliability tested	• Ambiguous items • Physician-level reliability not estimated
CAHPS 2.0	• Strong development and testing of reliability and validity • Estimates physician-level reliability • Includes three E&E items	
PEI	• Rigorous testing	• Limited scope

E&E = engagement and enablement

Endnotes

[i] Hutchinson et al (1999) surveyed directors of public health and complaints managers of all UK Health Authorities and boards, Local Medical Committee (LMC) secretaries, Community Health Council chief officers, and medical directors and complaints managers of a sample of NHS hospital trusts.

[ii] Doctors' and patients' assessments do sometimes concur. A study by Greco et al (2002) demonstrated high correlations between patients' and GP examiners' judgements of GP registrars' interpersonal skills.

'[iii] **Doctor-level reliability'** (an estimate of the number of observations required to achieve a given level of reliability) has to be calculated for each individual questionnaire and depends on 3 things: 1) the measurement characteristics of the questionnaire itself, 2) the level of confidence required of the results – fewer responses are needed for a process which needs to achieve reliability of 0.7 (which might be fine for formative purposes) than for reliability of 0.8 (eg for high-stakes summative assessments) and, 3) the number of levels of doctors' performance which are to be distinguished (fewer patients are needed to separate the best 60% from the worst 40% than to distinguish five levels of performance, i.e. the top 20% from the next best 20% and the middle 20% and so on). Crossley et al (2005) used generalisability theory to predict how many observations are required to achieve a given level of reliability. They found 15 ratings pooled were 70% representative of the views of adults about a doctor. (Violato et al 2003) found 25 patients were required to achieve generalisability of 0.70. Similarly, Webster (1989) estimated that a sample of 25 patients was sufficient to provide a reliable indicator of a doctor's performance. However, Nelson et al (2004) found the number of patients needed to achieve reliability of 0.80 at the clinician level was 66 for an 11-item scale (or 77 for primary care physicians only). Hays et al (2003) cite three studies which, collectively, estimate that 20-24 patient responses are necessary to obtain precise enough information for comparisons between physicians (Nelson et al 2006, Merterko et al 1994, Swanson et al 1990)

[iv] Since 1998 the NHS has, through the National Patient Survey programme, co-ordinated by the Picker Institute on behalf of the Healthcare Commission, obtained feedback on patients' experience and satisfaction in large-scale questionnaire surveys.

[v] There is no consensus on the precise meaning and definition of empathy, but it has been suggested that in the clinical context, empathy involves an ability to (i) understand the patient's situation, perspective and feelings (and their attached meanings); (ii) to communicate that understanding and check its accuracy; and (iii) to act on that understanding with the patient in a helpful (therapeutic) way.

[vi] **Internal consistency** assesses the extent to which the items relating to a particular dimension in a scale tap only this dimension and no other. The most commonly used method for assessing the internal consistency of patient questionnaires is Cronbach's alpha. A low alpha coefficient indicates that the items do not belong to the same conceptual domain (Bowling 2002).

Factor structure refers to the underlying dimensions of an instrument. Factor analysis can be used to define dimensions, each of which contains items which group together in a consistent way, to form a manageable set of variables.

Doctor-level reliability is a prediction of how many observations are required with different test formats to achieve a given level of reliability. Generalisability theory can be used to estimate how

many patient responses are required in order to assess a doctor's performance within acceptable limits of confidence. Adding items and adding observers will both increase generalisability (Violato et al 2003)

Conventionally, a reliability coefficient of 0.8 is desirable for 'high stakes' assessments such as certification procedures, although a lower reliability may be acceptable for other purposes (Davies and Howells 2004).

[vii] **Face** validity is an intuitive judgment about the relevance, reasonableness and clarity of the items within a questionnaire.

Criterion validity is a measure of the correlation between the results from the instrument with another measure (or criterion) that is itself accepted as valid (Baker & Whitfield 1992b). Two aspects of criterion validity are concurrent and predictive validity. Concurrent validity corroborates independently that the instrument measures what it intends to (e.g. against observable criteria). Predictive validity is demonstrated if the instrument predicts future changes in key variables in expected directions (Bowling 2002)

Construct validity is corroboration, against other measures, that the instrument measures the underlying concept it purports to measure (Bowling 2002). If an instrument has 'construct' validity, it will produce results that would be predicted by existing research or theory. Suppose, for example, existing research shows that patients of doctors with good communication skills have better outcomes than patients of doctors with poor communication skills. If an instrument records better communication skills in the doctors whose patients have good outcomes than in those whose patients have poor outcomes, this finding could be said to support the instrument's construct validity (Davies and Howells, R 2004).

[viii] MEDLINE, AMED, CINAHL and Psychinfo were searched by questionnaire name and author name for studies relating to the development and testing of the identified questionnaires for validity and reliability. The following combinations of search terms were used:

Topic	Author
General Practice Assessment Questionnaire OR GPAQ OR General Practice Assessment Survey OR GPAS	Ramsay-J$ OR Campbell-JL$ OR Schroter-S$ OR Green-J$ OR Roland-M$
Doctors$ Interpersonal Skills Questionnaire OR DISQ	Brownlea-A$ OR Buckley-B$ OR Field-D$ OR Francis-W$ OR Greco-M$ OR McGovern-J OR Cavanagh-M OR Pocklington-S$ OR Spike-N$ OR Powell-R$ OR Sweeny-K$ OR Broomhall-J$ OR Beasley-P
Consultation Satisfaction Questionnaire OR CSQ	Baker-R$ OR Whitfield-M$
CARE Measure	Mercer-S$ OR Maxwell-M$ OR Heaney-D$ OR Watt-G$ OR McConnachie-A$
SHEFFPAT	Crossley-J$ OR Davies-H$ OR Eiser-C$
ABIM AND patient questionnaire	Lipner-R$ OR Blank-L$ OR Leas-B$ OR Fortna-G$

(Physician Achievement Review OR PAR) AND questionnaire$	Hall-W$ OR Violato-C$ OR Lewkonia-R$ OR Lockyer-J$ OR Fidler-H$ OR Toews-J$ OR Jennett-P$ OR Donoff-M$ OR Moores-D$ OR Marini-A$
Patient Satisfaction Questionnaire OR PSQ	Marshall-G$ OR Hays-R$ OR Sherbourne-C$ OR Wells-K$ OR Ware-J$ OR Snyder-N$ OR Wright-W$ OR Davies-A$
CAHPS 2.0	Agency for Healthcare Research and Quality OR AHRQ OR Hargraves-J$ OR Hays-R$ OR Cleary-P$
Patient Enablement Instrument OR PEI	Howie-J$ OR Heaney-D$ OR Maxwell-M$ OR Walker-J$ OR Freeman-G$

[ix] Two questionnaires were, unfortunately, not included. The Communication Assessment Tool was for use by all member boards of the American Board of Medical Specialties (ABMS) for use as part of maintenance of certification. However, insufficient information about the instrument, its development processes and its validation was available to evaluate it.

An instrument is being developed on behalf of the General Medical Council (GMC) for use potentially in the revalidation process. However, its development is not yet complete, so we were unable to include it in this review.

[x] GPAQ is shortened version of General Practice Assessment Survey (GPAS), which was originally based on Primary Care Assessment Survey (PCAS), developed in USA by (Safran et al 1998). PCAS items were based not on primary research with patients, but on the Institute of Medicine's 1994 formal definition of primary care which emphasised the sustained partnership between patient and primary care physician, and that primary care occurs in the context of the family and the community (Institute of Medicine 1994).

General Practice Assessment Survey (GPAS) was modified from PCAS for use in UK to reflect differences in primary care provision between UK and USA, and to reflect five aspects of care highly valued by patients as identified in the literature: availability and accessibility; technical competence; communication skills; interpersonal attributes; organisation of care (Ramsay et al 2000d, Wensing M et al 1998, Buetow 1995, Smith and Armstrong, D 1989) Some items were reworded or removed and new ones added, including a multiple-item scale relating to technical competence (Ramsay et al 2000b, Ramsay et al 2000c). The 'acceptability to patients' of GPAS was tested by analysing item non-response, and further qualitative interviews were carried out to explore patients' understanding of issues such as technical care (Bower et al 2002b).

[xi] The Medicare Managed Care version of CAHPS (MMC-CAHPS) has been used to survey Medicare beneficiaries in managed care health plans annually since 1997 (Zaslavsky & Cleary 2002, Zaslavsky et al 2000).

[xiii] The items are based on the section in *Good Medical Practice* on 'communication with patients' so is likely to be suitable for revalidation and in-training assesment (Crossley et al, 2005)

[xiv] The attributes measured by the PAR General Practitioner patient questionnaire are: Patient Interaction (the physician listens, answers questions and demonstrates interest, empathy and respect for the patient during an examination period. Patients indicate whether they would return to or refer a friend to the physician); phone communication (the availability of a doctor by phone

after hours for urgent medical problems); personal communication (the physician adequately explains illness/injury, preventative measures, treatment options, and medication regimen and side effects); information for patients (the physician provides proper information regarding medical problems, return appointments, reporting of test results, referrals to specialists, tracking of prescription and non-prescription medication and patient education); office staff (the staff is pleasant, helpful, capable, professional and able to maintain confidentiality); physical office (the office is accessible, clean, private, and appropriately sized); appointments (appointments can be made quickly and wait time for scheduled appointments is not excessive).

[xv] a) The GPAQ covers four pages, including 19 (post-consultation version) or 20 (postal version) items with many sub-items, including three free text questions. The consultation version of the GPAQ includes three items on 'enablement' adapted from the Patient Enablement Instrument (see 5.10), asking *"after seeing the doctor today, do you feel…*

Able to understand and cope with problem(s) or illness

Able to cope with your problem(s) or illness

Able to keep yourself healthy

with response options: *"much more than before the visit," "a little more than before the visit," "the same or less than before the visit"* and *"does not apply."* The authors report that these items were less well understood in a postal context (National Primary Care Research and Development Centre (NPCRDC) 2004) so they were included only in the post-consultation version.

b) DISQ contains 12 items on one page, and covers the following topics: Greeting; listening; explanation; eliciting concerns; reassurance; time; consideration of personal situation; respect; privacy and dignity; recommendation to friends.

DISQ asks patients to rate the doctor's performance on a five-point scale from *"poor"* to *"excellent"* on these areas:

"The warmth of this doctor's greeting to me was.."

"The respect shown to me by this doctor was…"

"The doctor's concern for me as a person in this visit was…"

DISQ includes two items which assess the doctor's capacity to ask about and listen to the patient:

"On this visit I would rate the doctor's ability to really listen to me as…"

"The opportunity the doctor gave me to express my concerns or fears was…"

DISQ asks patients to rate whether the doctor takes a holistic approach to the consultation:

"This doctor's consideration of my personal situation in deciding a treatment or advising me was…"

DISQ includes a well-constructed single global satisfaction item, answered on a five-point likert scale from *"poor"* to *"excellent."*

"The recommendation I would give to my friends about this doctor would be.."

c) The CARE instrument consists of ten items on one page, which ask patients to rate doctors on a 5-point likert scale from *"poor"* to *"excellent,"* asking for each *"How was the doctor at.."*

Making you feel at ease; Letting you tell your 'story'; Really listening; being interested in you as a whole person; fully understanding your concerns; showing care and compassion; being positive; explaining things clearly; helping you take control; making a plan of action with you.

The last two items are related to enablement.

It also asks how important the items were in the context of this consultation. This is a good question, but the value of asking it once in relation to all the questions is dubious. It may be more useful to ask it once for each item, although this would have space implications.

d) The two-page PSQ-18 retains many characteristics of the PSQ III. Its 18 items measure each of the seven dimensions of satisfaction with medical care measured by the PSQ III (Marshall and Hays, RD 1994): general satisfaction; technical quality; interpersonal manner; communication; financial aspects; time spent with doctor; accessibility and convenience.

e) The PEI questionnaire includes only six items on one page. It focuses specifically on outcomes related to themes of patient-centredness and empowerment: issues related to, but conceptually distinct from, patient satisfaction (Howie et al 2000, Howie et al 1998). The authors define 'patient-centredness' as an indication by doctors of a commitment to value patients' contribution to deciding what is wrong with them and how their care should be managed. Empowerment is taken to mean that patients are helped to understand the nature of their problems and enabled to manage their own illness. Response options for the PEI are *"much better"/"Much more," "better"/"more"* or *"same or less,"* and the items ask, *"As a result of your visit to the doctor today, do you feel you are..." "able to cope with life", "able to understand your illness", "able to cope with your illness"; "able to keep yourself healthy"; "confident about your health, ."able to help yourself".*

[xvi] The development process of the ABIM questionnaire used in the study by Lipner et al (2002) is not described in the published literature. This paper refers to a paper by Weaver et al (1993) but this appears to describe the development of an instrument which differs significantly from the one used by Lipner et al.

[xvii] Sometimes questions deliberately omit 'don't know' options in order to force an answer one way or the other. But this would not be appropriate for our questionnaires if the intention is to get as close as possible to patients' views.

[xviii] The GPAQ 'manual' includes some guidance but it is not very stringent.