THE USE OF PATIENT-SATISFACTION SURVEYS AND OF MULTI-SOURCE FEEDBACK TOOLS IN APPRAISAL AND IN REVALIDATION

Hard-working grass roots GPs such as me will want further reassurances about Patient Satisfaction Surveys and MSF tools if we are expected to use them in appraisal and for revalidation. These are my personal views and reflections on these tools.

Key points

- A well-designed and scientific patient satisfaction survey or multi-source feedback (MSF) tool, giving timely feedback to the doctor, should be helpful for his personal reflection and for discussion at appraisal
- Neither of the two patient satisfaction tools currently used for the Quality and Outcomes Framework (QOF) has been formally assessed for reliability and their validation has been sub-optimal
- Currently used Patient Satisfaction Surveys are subjective and subject to huge elements of bias and to many variables outside the doctor’s control
- They are therefore unethical
- Some studies have shown no benefit, and even adverse results, from the use of MSFs
- Patient-satisfaction surveys and MSF tools are not sufficiently robust for revalidation
- A large study of a patient satisfaction survey used in Australia showed that it did not help GPs to improve patient-satisfaction over a nine year period
- Any tool used must be useful to GPs, helping GPs to improve their practice
- Any tool used must be acceptable to all GPs using it and GPs must have confidence in it
- Qualitative feedback is an essential part of any survey but commercial companies are not qualified to interpret it
- Qualitative feedback should be given to the GP at the end of each day for his own personal reflection and for later discussion at appraisal
- The GP concerned (perhaps with help from an appraisal discussion) is the only person qualified to interpret, and to reflect on, the results of Patient-Satisfaction Surveys and of MSF tools regarding himself.
- Different tools may be needed for regular members of a practice and for locum GPs
Introduction

Patient Satisfaction Surveys

Whilst the tradition of ‘listening to the patients’ is almost as long as the NHS itself, the prominence given to the patient satisfaction survey can be traced back to the Griffiths report, which encouraged the use of market research to obtain consumers’ views.iii

Purchasing authorities have been urged to pay more heed to locally expressed views of the quality of the service since the early 1990’s.iii It has also been recognised for about the same length of time that in judging the quality of hospital services, the judgements of patients alongside their clinicians is an intrinsic part of the quality measurement process.iv

Patient satisfaction surveys have been used widely in General Practice for many years and their use as a market research tool in general practice was researched and highlighted in 1994.v

They have also been used under the new contract as part of the Quality and Outcomes Framework (QOF) and GPs have received reimbursement for conducting surveys. The new RCGP “Guide to Revalidation for General Practitioners”vi states that such surveys will be used as a tool in revalidation.

Such surveys cover many aspects of a patient’s experience in using a practice, from the experience of making telephone calls to reception to evaluating aspects of the consultation.

The RCGP ‘has commissioned a review of patient surveys and will recommend which ones are appropriate for use in revalidation…to seek the views of the patients actually consulting the GP.’vii This work is being done by the same team that has already reported on tools for use in MSF.

‘The most important aspect of undertaking patient surveys is the reflection upon the results and, if appropriate, implementing change.’viii RCGP “Guide to the revalidation of General Practitioners”, p16.

This paper is partly based on my experience of using the General Practice Assessment Questionnaire (GPAQ)ix as a patient satisfaction survey on three separate occasions as a salaried GP (with the results analysed by the practice manager on one occasion and by independent companies on two other occasions) and on my use of GPAQ on one occasion as a locum GP (when I analysed the results myself).

It was developed at the National Primary Care Research and Development Centre at The University of Manchester for the 2003 GP contract.
Multi-Source Feedback

“The RCGP Guide to the Revalidation of General Practitioners”\(^v\) states that, in addition, “when revalidation is fully established, each GP will be required to submit evidence from two Multi-Source Feedback exercises (MSF)s, one undertaken in the first 2 years of the 5-year revalidation period and one in the last 2 years”.

It also states that “The RCGP has commissioned a review of MSF instruments and will recommend which ones are appropriate for use in revalidation. At present only the GMC’s MSF instrument meets RCGP requirements”. This is as a result of the final report “Comparison of Multisource Feedback Instruments Designed for GP’s in the UK”, prepared by Jocelyn Lockyer and Herta Fidler at the University of Calgary. This is a very thorough piece of scientific work but I assume that neither of these people has any experience of clinical work in British General Practice.

The German Military first began gathering feedback from multiple sources in order to evaluate performance during World War II\(^xi\). Also during this time period, others explored the use of multi-rater feedback via the concept of T-groups.

One of the earliest recorded uses of surveys to gather information about employees occurred in the 1950s at Esso Research and Engineering Company\(^xii\). From there, the idea of 360-degree feedback gained momentum, and by the 1990s most Human Resources and Organization Development professionals understood the concept. The problem was that collecting and collating the feedback demanded a paper-based effort including either complex manual calculations or lengthy delays. The first led to despair on the part of practitioners; the second to a gradual erosion of commitment by recipients\(^xiii\).

Multi-rater feedback use steadily increased in popularity, due largely to the use of the internet in conducting web-based surveys\(^xiv\). Today, studies suggest that over one-third of U.S. companies use some type of multisource feedback\(^xv\). Others claim that this estimate is closer to 90% of all Fortune 500 firms (Edwards & Ewen, 1996). A google search for multi-source feedback on UK websites demonstrated that it was widely used in the NHS, and found a few references stating that it was also widely used outside the NHS in the UK\(^xvi\).

A study on the patterns of rater accuracy shows that length of time that a rater has known the person being rated has the most significant effect on the accuracy of a 360-degree review. The study shows that subjects in the group “known for one to three years” are the most accurate, followed by “known for less than one year,” followed by “known for three to five years” and the least accurate being “known for more than five years.” The study concludes that the most accurate ratings come from knowing the person long enough to get past first impressions, but not so long as to begin to generalise favourably\(^xvii\).

It has been suggested that multi-rater assessments often generate conflicting opinions, and that there may be no way to determine whose feedback is accurate\(^xviii\). Studies have also indicated that self-ratings are generally significantly higher than the ratings of others (Lublin, 1994; Yammarino & Atwater, 1993\(^xix\); Nowack, 1992).
Some concerns about Patient Satisfaction Surveys

Williams cautions against making “assumptions about the nature and meaning of expressions of 'satisfaction'”. He writes:

“Patient satisfaction is now deemed an important outcome measure for health services; however, this professed utility rests on a number of implicit assumptions about the nature and meaning of expressions of 'satisfaction'. Through a review of past research findings this paper suggests that patients may have a complex set of important and relevant beliefs which cannot be embodied in terms of expressions of satisfaction. Consequently, many satisfaction surveys provide only an illusion of consumerism producing results which tend only to endorse the status quo. For service providers to meaningfully ascertain the experience and perceptions of patients and the community then research must first be conducted to identify the ways and terms in which those patients perceive and evaluate that service.”

Most hard-working grass roots GPs would want to know:

What do we want to achieve with patient satisfaction surveys and MSFs? Is it to give doctors helpful feedback, for reflection and for discussion in appraisal, or is it to prove fitness to practice for revalidation? Is it appropriate for them to be a part of revalidation at all? Even if a tool is reliable and validated in a scientific sense, it doesn't necessarily mean that the results matter, that they are useful or that they are an appropriate and sufficiently robust tool to be used in revalidation. Its use in revalidation must be evidence-based. If revalidation is about fitness to practice, then patient satisfaction surveys and MSFs may not be appropriate, since revalidation is a pass/fail assessment and since, as this paper will demonstrate, these tools are not sufficiently robust to tell us about fitness to practice. If we want to improve our patient-doctor-relationships, then they would be suitable only for appraisal, which is part of revalidation anyway.

How do we achieve our goals? What sort of questionnaires do we need? Different questionnaires for GPs working regularly in a practice and for locum GPs may be needed.

Is the use of patient satisfaction surveys and of MSFs in appraisal and/or revalidation acceptable to all GPs, and are the questionnaires to be used acceptable to all GPs?

In a chapter entitled “Getting Evidence into Practice” Trisha Greenhalgh writes about the “human factor”. She says:

“People are not passive recipients of innovations. Rather (and to a greater or lesser extent in different individuals), they seek innovations out, experiment with them, evaluate them, find (or fail to find) meaning in them, develop feelings (positive or negative) about them, challenge them, worry about them, complain about them, “work round” them, talk to others about them, develop know-how about them, modify them to fit particular tasks, and attempt to improve or redesign them.”

Are the results from Patient Satisfaction Surveys and MSFs useful, are the results reproducible in other conditions (are they reliable) and do they measure what they claim
to measure (are they valid)? Are they useful to me and my patients? Will they improve care? Are they worth doing?
Is the questionnaire well designed and is the whole process scientifically sound?

Hart \textsuperscript{xxii} cautions against making assumptions about the validity of questionnaires. He writes:

“…there are severe doubts whether such traditional methods measure anything other than 'hotel services' and their construction reflects the interests of the producers rather than the consumers of healthcare.”

Hart continues:

“There is some concern, expressed cogently by Carr-Hill\textsuperscript{xxiii} after his review of some 300 patient satisfaction surveys that the majority of them are producer-led:

“Once the fieldwork is over, there is considerable temptation to forget that what are confidently described as respondents' views are only their replies to questions devised by the researcher and not necessarily the patients' own views and priorities. Thus it is commonplace to observe that health service policy has been steered by providers' perceptions and definitions of good practice.”

“Carr-Hill is also concerned with the many methodological inadequacies which he details as a result of his survey. These range from problems with the framing of the questions, the avoidance of evaluation of clinical practice, the inadequate ways in which samples relate to the populations from which they are drawn and the cavalier treatment of non-response rates. To this, we may add the fact that many patient surveys appear to be exhibit a halo effect in which satisfaction rates seem to be uniformly high at over 80%, perhaps reflecting a reluctance to criticise nurses\textsuperscript{xxiv}. There are indications, however, that much more attention is now being paid to questionnaires in terms of both their construct validity\textsuperscript{xxv} The timing and location of the survey may itself be a critical factor. In a study of particular relevance to a concern with outpatients\textsuperscript{xxvi}, it is shown that there is a clear decay in satisfaction levels when patients are interviewed at home rather than in the outpatient clinic. But probably the greatest single source of dissatisfaction with the traditional survey is its superficiality. The most common method of data collection involves the use of pre-coded self-completion questionnaires\textsuperscript{xxvii}

But as Rigge\textsuperscript{xxviii} has pointed out:

“Handing out tick-in-the-box patient satisfaction questionnaires and then sitting smugly back if the results indicate that most patients are satisfied with the service they have received (as many such quantitative methods do) is no substitute for genuine consultation.”

Is the use of Patient Satisfaction Surveys and of MSFs \textit{fair and ethical}? If it seen as unscientific, and a waste of time and money (especially if it has to be funded out of doctors’ own pockets) it is unlikely to be a useful tool in improving performance and may even cause distress and add to burnout. It would then clearly be unethical.
I believe that a patient satisfaction survey should give immediate feedback to a doctor, who knows all the circumstances behind a particular consultation, for his personal reflection and to fill in his Johari Window. The qualitative results of such surveys might help a doctor in his professional development and can be discussed at an appraisal.

There have been suggestions that, for revalidation, locums may have to pay for the cost of the surveys themselves. This would involve the patient returning the survey to a company in a freepost envelope, and in that company then collating, analyzing and distributing the results. The cost of this is likely to be prohibitive for locum GPs to fund out of pocket.

The idea behind this bureaucracy is no doubt to ensure that the GP does not try to add, modify or delete forms from the survey. The question then becomes: do the currently used surveys provide results of such reliability, validity, usefulness and importance to justify the considerable expense and bureaucracy involved? In my opinion, the answer to that is a resounding NO, for reasons explained below. I strongly believe that a large number of doctors do not find the exercise either helpful or scientific.

Problems relating to feedback on locum GPs have been described in an article on the website of the National Association of Sessional GPs (NASGP)xxix.
Do MSF exercises work?

Yes

Several studies (Hazucha et al., 1993; London & Wohlers, 1991; Walker & Smither, 1999) indicate that the use of 360-degree feedback helps people improve performance. In a 5-year study by Walker and Smither (1999), no improvement in overall ratings was found between the 1st and 2nd year, but higher scores were noted between 2nd and 3rd and 3rd and 4th years. A study by Reilly et al. (1996) found that performance increased between the 1st and 2nd administrations, and sustained this improvement 2 years later. Additional studies show that 360 feedback may be predictive of future performance (Maylett & Riboldi, 2007).

No

However, others indicate that the use of multi-rater assessment may not improve company performance. A 2001 Watson Wyatt study found that “Multisource feedback is gaining in popularity, but is linked to a 10.6 decrease in market value—most likely because it's a challenge to get this feedback right. It requires an open culture, well-trained participants and follow-up. When one or more of these elements is missing, multisource feedback can be a lengthy distraction that interferes with teamwork”. The study notes that while nothing is inherently wrong with these practices, many organisations implement them in misguided ways. "These types of complex, process-driven programs can actually destroy value if they aren’t aligned with strategy and executed properly. For example, implementing multi-source feedback in a company that does not operate in an open environment will likely yield negative results, especially if the employer doesn’t follow up by acting on employee opinions.

John Sullivan, professor of HR management at San Francisco State University, states that "there is no data showing that [360-degree feedback] actually improves productivity, increases retention, decreases grievances, or is superior to forced ranking and standard performance appraisal systems. It sounds good, but there is no proof it works." Similarly, Siefert, Yukl, and McDonald (2003) state that there is little evidence that the multi-rater process results in change.
Use of Patient Satisfaction Surveys in Appraisal and revalidation

Patient surveys have been used in appraisal and have provided some information on how a patient rated a doctor on a particular day. They have been useful for a personal discussion with an appraiser and few would argue with that. However, what is proposed now is very different. It is proposed that they will be used as a component in revalidation, which is a pass/fail assessment. Any tool used in this way MUST be totally robust and scientific. If it is not, it is unethical and unacceptable.

Appraisal and revalidation came out of the Shipman inquiry, and the recommendations were designed to ensure that this tragedy could never happen again. Appraisal and revalidation will not achieve that; indeed, as has often been commented before, Shipman would probably have sailed through both appraisal and revalidation since he was meticulous in record keeping and was popular with his patients. He would probably have scored highly in Patient Satisfaction surveys, whereas some excellent doctors today could fail revalidation if MSFs and Patient-Satisfaction Surveys are not fit for purpose. There is a real risk that the results of a survey depend very much on factors on the day of the survey outside the doctors control, which would mean that the same results are not reproducible on another day or in another practice (the results are therefore not reliable), that these surveys do not measure what they claim to measure (they are therefore not valid) and that the results are over-influenced by qualities such as friendliness and likeability; these have little, if anything, to do with fitness to practice. Sometimes we have to disappoint our patients in order to be good doctors.

Reliability and validity
These two terms are helpfully explained in this paper by Fitzpatrick xxxvi:

Reliability
Fitzpatrick writes:

“Reliability is concerned with the extent to which a questionnaire produces the same results on separate occasions of use (my underlining). Clearly, examining such features is not easy, given that there may be low agreement between two administrations of a questionnaire, which may be due to real changes in patients’ views. Remarkably few studies have examined what is formally known as the "test-retest" reliability of questionnaires of patient satisfaction, but the results have been encouraging. 'An alternative approach is by examining "internal reliability." One common form of this is "split half" reliability, which examines the extent of agreement between the halves of a questionnaire or section of a questionnaire considered to be measuring a particular dimension of satisfaction. Again, results of such examinations are satisfactory.”
**Validity**

Fitzpatrick continues:

“Validity is much more difficult to examine and, compared with reliability, is beyond the means of most simple surveys to evaluate in any real sense. This is because validity of a questionnaire is the requirement for it to measure what it claims to measure (my underlining) and it is difficult to imagine any ultimate gold standard against which to assess a questionnaire of satisfaction. However, some elaborate studies showed that such questionnaires can relate to other measures in theoretically expected ways (so called "construct validity"). Patients' views about consultations as expressed in questionnaires correlate with independent measures of doctors' interpersonal skills, communication styles, and technical proficiency”.

The producers of GPAQ claim that the questions in GPAQ (and in GPAS, an earlier version) have been “extensively tested and validated in British general practice” xxxvii.

How, exactly, can one validate these surveys in any meaningful sense of that word?

This article from the BJGP xxxviii demonstrates that, at the time the article was written, the reliability and validity of GPAQ itself had not even been properly evaluated, contrary to the claims of the producers at the University of Manchester:

The authors write: “No paper reported the reliability and validity of the GPAQ, but three papers assessed an earlier version (the GPAS). No published evidence could be found that the IPQ (Improving Practices Questionnaire, an alternative tool authorized for use in the QOF), GPAQ, or GPAS have been validated against external criteria. The GPAS was found to have acceptable reliability and test–retest reliability. Neither of the instruments mandated by the GMS contract has been formally assessed for reliability: their reproducibility remains unknown. The validation of the two questionnaires approved by the QOF to assess patient satisfaction with general practice appears to be suboptimal.”

I’m not convinced that GPAQ measures what it claims to measure (i.e. it is not valid) and, if the crude data obtained from the survey is “garbage” (unscientific, subjective, unreliable, not validated and subject to much bias and to multiple variables), no matter how rigorous the computer analysis is, the data being produced will also be “garbage”. There is a risk that results in Patient-Satisfaction surveys are unduly influenced by qualities such as friendliness and likeability. Shipman would probably have scored well on that count, but those qualities don’t tell us about fitness to practice.
Acceptability

It is not sufficient just to have a company reviewing the MSFs or surveys from a scientific point of view, even though that is important. The use of Patient Satisfaction Surveys as a tool in revalidation and/or appraisal must be acceptable to all GPs.

It may be necessary to use separate surveys/MSFs for regular doctors in a practice on the one hand and locums on the other. In that event, the survey/MSF must be acceptable to the group of GPs using it.

GPs must have confidence that the tool is useful in assessing and improving their performance and not just another tick box exercise without benefit to anybody. I don’t think many GPs I know have much confidence in GPAQ, even if someone were to tell them that it had been piloted and validated and was reliable. Moreover, even if a respected author writing in a good journal tells you that a survey is reliable and validated, it doesn’t necessarily mean that it is either of these things or that it is useful and meaningful. Intuition tells me that GPAQ is subjective and subject to huge elements of bias and variables.

Any tool that is used in revalidation must also be useful to GPs, and not just a tick-box exercise or used to check that all of us get adequate marks. Experience in this very large Australian study showed that patient satisfaction did not improve over the high baseline. The authors comment: “It is likely that this and the presentation of results made it difficult for GPs to use the survey to improve their practices. A more useful survey would be more sensitive to detect negative patient opinions and provide integrated feedback to GPs”.

Workload

The MSF from the GMC contains 17 items and, although this is one of the shorter MSF tools available, it will still take much time to complete. The RCGP “Guide to the Revalidation of General Practitioners” states that “In uncomplicated cases the questionnaire should take 10 to 20 minutes to complete, but it may take longer if reflection and consideration are required”. If GPs and practice managers have to complete one of these MSF exercises twice every 5 years on all their doctor colleagues and regular locums, the workload may be considerable. Taken together with all the other requirements for enhanced appraisal and revalidation, we are certainly talking about a substantial increase in workload for GPs, many of whom are already at breaking point. This will reduce the time GPs can spend on their core job, that of looking after patients. If people’s time or resources are wasted on surveys that bring little or no benefit, and if extra workload is excessive, ultimately burnout could result.
What are the benefits and features of a well-designed patient-satisfaction survey?

A well-designed patient-satisfaction survey would give the patient an opportunity to say exactly what they like and dislike about a practice and about a particular doctor. For the purposes of appraisal and revalidation of locums we are talking about feedback about a particular doctor only.

The feedback to the doctor should be immediate and, if the patient consents, should identify the patient making the feedback so that the mature professional can reflect better on a particular consultation and ultimately provide a better service to his patients. The survey should be clear and easy to complete. It should not take too much time.

What are the problems with currently used patient-satisfaction questionnaires?

I have always received very good personal results in such surveys but I nevertheless believe that they are unscientific, subjective and, in some respects, poorly designed. They give no useful qualitative feedback to the doctor in a timely manner on how he could have improved his performance in a particular consultation.

In GPAQ, the questions in section 10 relate to the particular consultation. Patients giving higher marks on these questions should tick boxes on the right of the form, whereas in section 11, where the questions also relate to the consultation and care, the patient should tick boxes on the left of the page. I have found that some patients who gave me particularly high marks in section 10 appeared to give me low marks in section 11.

I suspect that patients were confused by the layout of the GPAQ form or did not read it correctly. If so, this would indicate a design flaw in the questionnaire. We should not be trying to catch patients out here, but to give them a simple and clear form to complete.

Can we measure everything?

General Practice, and particularly the consultation, is, at least in part, an art form and includes concepts such as caring. This is reflected in the motto of the Royal College of General Practitioners (RCGP), *cum scientia caritas* (“science with caring”).

Dr Roger Neighbour, former President of the RCGP, and an expert on consultation skills, often quotes John Heron on his consultation skills courses at the RCGP: “A professional has a wide range of options to select from, and can move cleanly and elegantly amongst them”.

These sorts of things are certainly not measurable and yet, in these patient satisfaction surveys, an attempt is being made to measure things that are not measurable. We should therefore question their validity.
Variables and bias

Patient Satisfaction Surveys
So many variables beyond the doctor’s control can influence the results of such surveys, patient scoring is so variable and subjective, and there are so many possible elements of bias that I believe such surveys are not sufficiently robust to be used for revalidation purposes.

Some authors have commented on the importance of recall bias in interpreting survey results. The authors of one paper write: “self-reported data is subject to respondent recall bias and may have affected the survey responses we received, especially from the elderly group”.

The following is a list of factors both in and outside the consultation that may influence the results of surveys. It is not exhaustive but gives just a few reasons why the process is subjective, subject to bias and unscientific:

- Age, gender and ethnic origin of both doctor and patient
- Language barriers
- Variable assessment of time spent. At least one patient gave me average marks for “time spent” when, in fact, I had given her 20 minutes in a 10 minute appointment slot. At least one other patient gave me an excellent rating for “time spent” for a 3 minute consultation, because his problem was dealt with completely in that time.
- The length of time a patient waited for a consultation. In my experience, patients are not good judges or reporters of how long they have had to wait. If they arrived some time before the appointment time, they may get a particularly inaccurate picture of how long they had to wait after the actual appointment time and may not check a watch before entering the room. This can affect not just their assessment of the time they had to wait on the form, but also the scores awarded for other criteria.
- Patients arriving late for appointments
- Patients asking the doctor to see a relative or friend as an “extra”
- Stress, workload and illness of the doctor
- Related to the point above, the number and complexity of the problems presented to the doctor, and the manner, timing and order in which they are presented
- Factors in the doctor’s family or other outside matters
- Relationships with colleagues and staff
- Patient expectations
- The time available to the doctor, usually 10 minutes but may be less in a busy on-call surgery
- How late the doctor is running and any other commitments
- Telephone interruptions, requests to sign scripts during surgery.
- The place where consulted, especially prisons, but also old and poor premises, poorly-equipped consulting rooms etc
- Computer/printer problems
- Case mix. A surgery of many viral infections, where the doctor gives only advice, does not give the antibiotic anticipated and does not have a chance to demonstrate his consultation skills at their best is likely to score less well than for a surgery
with a varied case mix. This scenario is extremely common in the winter months. A surgery with an above average number of mental health cases will cause the doctor to run late, may cause stress for the doctor and may influence scores in a huge variety of ways.

- Status of doctor: locum or regular practice member
- How well the doctor knows the patient
- Regular patient or infrequent attender
- Demeanour of receptionist
- Patient’s general feelings about practice, irrespective of particular doctor’s ability
- Did the patient decide to complain before they even entered the room (a common event, according to the MPS)?

**MSFs**

Similar concerns apply to MSFs. The following is a list of some of the factors that may unfairly affect the feedback someone receives:

- Contractual problems
- Different attitudes to pay, workload, ethics, employment issues, morality, religion and politics and different outlooks on life
- Personality clashes
- Someone with lower clinical standards giving feedback on someone with higher standards
- Problems in a personal relationship between two members of staff
- Feelings of exploitation

These concerns are particularly relevant when the profession is so deeply divided, especially between partners and salaried doctors. Salaried GPs often have different attitudes to contracts, pay and workload from their partner employers and often feel exploited.iii

**What to include in the questionnaire**

The subject of Patient Satisfaction Surveys is discussed widely in the literature, and advice has been given on designing questionnaires and conducting surveysxliv

A review article on “Selecting, designing, and developing your questionnaire” xlv by Boynton and Greenhalgh in the BMJ is particularly helpful. It comments on how useful open questions and free text can be, provided someone is skilled in interpreting the responses. Here are some important extracts from their article:

“Questionnaire items may be open or closed ended and be presented in various formats”.

“Closed ended designs enable researchers to produce aggregated data quickly, but the range of possible answers is set by the researchers not respondents, and the richness of potential responses is lower. Closed ended items often cause frustration, usually because researchers have not considered all potential responses."
Ticking a particular box, or even saying yes, no, or maybe can make respondents want to explain their answer, and such free text annotations may add richly to the quantitative data. You should consider inserting a free text box at the end of the questionnaire (or even after particular items or sections). Note that participants need instructions (perhaps with examples) on how to complete free text items in the same way as they do for closed questions. If you plan to use open ended questions or invite free text comments, you must plan in advance how you will analyse these data (drawing on the skills of a qualitative researcher if necessary). You must also build into the study design adequate time, skills, and resources for this analysis; otherwise you will waste participants’ and researchers’ time. If you do not have the time or expertise to analyse free text responses, do not invite any.”

“A valid questionnaire measures what it claims to measure. In reality, many fail to do this.”

“Reliable questionnaires yield consistent results from repeated samples and different researchers over time. Differences in results come from differences between participants, not from inconsistencies in how the items are understood or how different observers interpret the responses. A standardised questionnaire is one that is written and administered so all participants are asked the precisely the same questions in an identical format and responses recorded in a uniform manner. Standardising a measure increases its reliability.”

In my view, the survey must contain qualitative feedback. Experience in the very large study of the Australian PPP tool (ibid) showed that GP performance was very high initially and there was minimal improvement over time. I do not know what the result would be in the UK if a similar study were attempted here, but I would speculate that it would yield similar results. In any case, we should try and learn from the Australian experience and design surveys that can lead to real improvements over time, if that is possible when the baseline is so high anyway.

I’m always very frustrated when I get survey results. They tell me that broadly patients are very happy; I sometimes get some positive qualitative feedback on the back of the form, but never any information whatsoever on how I could get even better. If a patient gives me less than a top scoring I want to know what I could have done to get a top scoring.

GPAQ question 18 states: “We are interested in any other comments you may have. Please write them here

Is there anything particularly good about your health care? (free text box)

Is there anything that could be improved? (free text box)

Any other comments?” (free text box)
Qualitative information like that needs skilled interpretation but it is invaluable to a GP. Indeed, only the GP concerned is in a position to interpret it reliably, possibly with the support of an appraiser. It cannot be interpreted by an independent company who know nothing of the circumstances behind a consultation. The danger is in how the company will interpret and present any qualitative feedback, perhaps ultimately resulting in loss of livelihood for the doctor if that tool is used in revalidation (pass or fail).

Nevertheless, I believe that these qualitative questions (with free text answers) should be asked and the answers fed back immediately to the GP for his own reflection and for discussion at appraisal.

**Feedback to doctor**

With current surveys, the results are posted to an independent company for analysis. The results are sent to practices several weeks or even months later. By this time, the doctor will have little or no memory of the consultations, there is no way of linking particular marks with a particular consultation and the results are purely quantitative.

Survey results show an overall score for each of a number of criteria and a national percentile ranking. Sometimes, the personal comments made by patients at the end of the forms are not reported. The doctor may be left feeling bewildered or frustrated; there is nothing to tell him how he could improve or on what went wrong in a particular consultation. The exercise has been unhelpful, perhaps intimidating, maybe resented, an almost total waste of time and may have had undesirable psychological effects. Is this ethical?

**Interpretation**

Great care must be taken in interpreting the results of patient satisfaction surveys. GPAQ state that differences in scores of less than 10% are not significant and I would guess that error margins are enormous. Therefore, a doctor with a score in the lowest percentile may have a score that is of no statistically significant difference from a doctor much higher in the percentile ranking.

Are commercial companies sufficiently reliable to interpret survey results, especially any qualitative feedback? Revalidation is a pass or fail assessment, and not one to assess progress, like appraisal. If patient surveys are used as part of revalidation, and if the RCGP decides to continue with their plan of having survey forms returned to a company for analysis, then the actions of the company receiving these anonymous surveys may well result in some GPs failing revalidation. The surveys then become almost like a complaints procedure without input from the doctor (contrary to guidance for the new (from April 1) complaints procedures). Since the survey form is anonymous, there is no way of identifying a patient who uses the form to complain or to make unsubstantiated or pejorative remarks about a doctor. Therefore, a doctor cannot defend himself or explain that the patient doing the survey was refused methadone, refused to go to a drug misuse team or GPSI, or that a man on parole was refused a sick note because he was not sick and simply missed his parole, or that a lady with a personality disorder comes in every week with a list of unreasonable expectations, etc. Such a situation is not just unfair; it also puts GPs livelihood at risk and is unethical.
A proposal

It would be much more useful if the survey forms (and, at the very least, the qualitative part of them) could be fed back to the GP at the end of that same day. He could then reflect on them in a timely manner, develop and reinforce things that he does well and act on bad habits immediately.

I propose that the form relating to a particular doctor (as opposed to a practice as a whole) should include 2 simple questions which every patient can understand, perhaps worded as: what did I do well? What could I have done better? Even if the RCGP insists on having a more detailed survey, sent to a company for analyzing (and this paper has highlighted the many flaws in that approach) the qualitative part of the survey should be detached and handed to the doctor at the end of the day. It should not be sent to a commercial company who are not qualified to interpret it.

There should be boxes for patients to write free text comments in answer to these two questions. Many doctors may find the results from 1 to 5 rating scales fairly meaningless and subjective. By contrast, comments such as “you didn’t look at me”, “you kept interrupting me”, “you kept messing around with the computer” or “you didn’t give me an opportunity to talk about what really mattered to me” could give valuable insight about a recurrent deficiency in a doctor’s consultation skills, or they may just reflect that, on that particular day, the doctor was running very late, his daughter was ill yet again, the patient had presented with 6 problems, the computer had crashed or that the printer kept chewing up the paper.

Only the doctor concerned is in a position to reflect on the qualitative survey findings. As a responsible professional he will certainly want to do so. He can then discuss the results and his reflections on them in an appraisal. The appraiser would be expected to look at a selection of the qualitative parts of the forms received and the doctor would be expected to demonstrate that he had taken the exercise seriously, had reflected carefully on any issues raised and could give an account of any factors that may have affected how well a particular consultation went.

This method would give timely feedback on things that really matter in an unthreatening way.
**Ethics**
A very important quote from the article by Boynton and Greenhalgh (ibid) above reads:

“A study is unethical if it is scientifically unsound, causes undue offence or trauma, breaches confidentiality, or wastes people’s time or money.”

I have demonstrated that the GPAQ survey was not properly evaluated for reliability and validity before being widely used in the QOF, that it is subjective and subject to bias and to many variables outside the doctor’s control. Any results from it are therefore unreliable. All this means that the GPAQ survey is scientifically unsound. Moreover, its widespread use in the QOF over a number of years has cost a lot of money and has involved a lot of people’s time. All this is wasted since the results are unscientific, unreliable and therefore useless. The GPAQ survey is therefore unethical, according to this review by two respected authors.

Does the enforced use of a survey by doctors who have no confidence in it cause psychological trauma? If so, it is unethical.

Is there a risk that the livelihood of good doctors is removed because the commercial company draws the wrong conclusions from the survey results? If so, it is unethical.

If GPs are asked to pay for conducting surveys but it is a waste of their time (and patients’ time) and money because no useful information to improve practice emerges, then the survey is unethical. It risks causing frustration, psychological trauma and the very livelihood of the doctor concerned.

**Conclusion**

It is inappropriate and unethical to expect GPs to use some current patient-satisfaction surveys in revalidation. These surveys purport to be scientific and to measure things in a useful way but they have not, in reality, been properly evaluated for reliability and validity.

Survey results should result in an improvement in practice but there is some evidence that some surveys do not achieve this aim. Surveys must be better-designed, with greater sensitivity for negative feelings of patients, so that they can result in improved care.

GPs should not be asked to spend large sums of money on having patient feedback forms analysed by a commercial company, only to receive an unreliable analysis, and impersonal, untimely and unhelpful feedback.

The quantitative results from current surveys are certainly not robust enough as a revalidation tool, even if the doctor has been prevented from adding, modifying or deleting forms.

My proposal is designed to provide doctors with an informal, unthreatening and simple tool that gives them rapid feedback on their consultations with patients, an opportunity to reflect on the results in a timely manner, to improve the quality of care they provide in a more informed way, and a basis for a simple, informal and constructive discussion in the appraisal.
**Taking this matter forward**

The scientific work so far commissioned by the RCGP has been done by a team in Calgary, Canada and I assume that neither of those people has worked in British General Practice. I feel that further work needs to be done by someone with extensive knowledge and experience of both the clinical aspects in British General Practice and of research methods. In the light of the concerns raised in this document, it may be worth the RCGP commissioning a report from an independent body with expertise in both the clinical and research domains in British General Practice, to cover the scientific and clinical aspects of the following:

- a review of the literature on Patient Satisfaction Surveys and MSF so that the process is evidence-based
- the benefits and disadvantages of patient-satisfaction surveys and of MSF
- an opinion on whether the results of such surveys/MSF exercises matter
- the concerns raised in this paper, especially sources of bias and variables outside the doctor’s control that could affect the results
- the measurability of the different criteria
- the scientific robustness, reliability and validity of the Patient Satisfaction Surveys and MSFs currently available,
- their appropriateness or otherwise as a tool in a) appraisal and b) revalidation
- their acceptability, confidence-inspiring and usefulness to two groups of GPs: regular practice doctors and locums
- what to include in the questionnaires
- the workload generated
- the interpretation of results
- the timeliness of feedback to doctors
- the use of open questions so as to generate maximum information for the doctor
- the ethical dimensions of performing Patient-Satisfaction Surveys and MSFs in a) appraisal and b) revalidation
- advising on the appropriateness of the use of Patient Satisfaction Surveys and of MSFs by locum GPs

Dr John Pike 30.4.09

I wish to thank Dr Una Coales and my wife, Dr Stephanie Franz, for their help in preparing this paper.
References


iii NHSME (1992), *Local Voices - The Views of Local People in Purchasing for Health*, EL(92)1, January


ix Details of the GPAQ survey may be found here: http://www.gpaq.info/


xiii http://en.wikipedia.org/wiki/360-degree_feedback


xvi http://www.google.co.uk/search?q=multisource+feedback&hl=en&cr=countryUK%7CcountryGB&start=60&sa=N


xliii The next 13 links all relate to the alleged exploitation of salaried GPs by partners:

http://www.pulsetoday.co.uk/story.asp?sectioncode=23&storycode=4122292
http://www.pulsetoday.co.uk/story.asp?sectioncode=39&storycode=4122222
http://www.pulsetoday.co.uk/story.asp?sectioncode=20&storycode=4121481
http://www.pulsetoday.co.uk/story.asp?sectioncode=20&storycode=4116882
http://www.pulsetoday.co.uk/story.asp?sectioncode=20&storycode=4116878
http://www.pulsetoday.co.uk/story.asp?sectioncode=39&storycode=4116864
http://www.pulsetoday.co.uk/story.asp?sectioncode=23&storycode=4116856
http://www.pulsetoday.co.uk/story.asp?sectioncode=39&storycode=4116638
http://www.pulsetoday.co.uk/story.asp?sectioncode=23&storycode=4116457
http://www.pulsetoday.co.uk/story.asp?sectioncode=39&storycode=4116325
http://www.pulsetoday.co.uk/story.asp?sectioncode=23&storycode=4116457
http://www.pulsetoday.co.uk/story.asp?sectioncode=23&storycode=4115171
http://www.pulsetoday.co.uk/story.asp?sectioncode=23&storycode=4115156


and


and

xlv Petra M Boynton and Trisha Greenhalgh. “Selecting, designing, and developing your questionnaire” British Medical Journal 2004;328;1312-1315, available at: http://www.bmj.com/cgi/reprint/328/7451/1312

xlvi http://www.gpaq.info/benchmarks%20consultation%202005_6.htm