

Measuring patient satisfaction for the Quality and Outcomes Framework

Matthew Hankins, Alice Fraser, Andrew Hodson, Claire Hooley and Helen Smith

ABSTRACT

The general medical services (GMS) contract Quality and Outcomes Framework (QOF) awards up to 70 points for measuring patient satisfaction with either the Improving Practices Questionnaire (IPQ) or the General Practice Assessment Questionnaire (GPAQ). The usefulness of data collected depends crucially on the validity and reliability of the measurement instrument. The literature was reviewed to assess the validity and reliability of these questionnaires. The literature was searched for peer-review publications that assessed the reliability and validity of the IPQ and GPAQ, using online literature databases and hand-searching of references up to June 2006. One paper claimed to assess the validity and reliability of the IPQ. 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, 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. It is recommended that future patient experience surveys are piloted for validity and reliability before being implemented widely.

Keywords

patient survey; questionnaire; Quality and Outcomes Framework.

INTRODUCTION

The Quality and Outcomes Framework (QOF) was introduced in 2004, under the new general medical services (GMS) contract,¹ to enhance patient care. This pay-for-performance programme includes quality indicators relating to clinical care, organisation of care, and patient experience. General practices have been able to meet the contract requirement for patient surveys by using one of two currently approved instruments: the Improving Practices Questionnaire (IPQ),² and the General Practice Assessment Questionnaire (GPAQ).³ Measuring patient satisfaction in this way and reflecting upon the results earns up to 70 points within the QOF.

Measurement of patient satisfaction is by no means straightforward. Unless carefully designed, questionnaires could introduce a positive bias,⁴ or reflect the responder's desire to please,⁵ rather than the extent to which they are truly satisfied. Before questionnaires are promoted for widespread use they require formal testing to demonstrate their ability to measure what they purport to measure (validity) and the precision, or reproducibility, of that measurement (reliability). The process of assessing the validity and reliability of questionnaires has been well documented.^{6,7}

While not straightforward, the design of a reliable and valid questionnaire is a piece of empirical work that should precede its adoption for research or survey purposes. A questionnaire should be validated and indices of reliability calculated as part of the early fieldwork. As with all empirical work, the gold standard for dissemination of the reliability and validity assessment is publication in peer-reviewed journals. A recent review of studies of patient satisfaction, however, found very little attention given to the reporting of reliability and validity.¹⁰ Without such data, the quality of research and validity of the conclusions cannot be properly assessed. This study, therefore, sought to critically review the empirical evidence for the reliability and validity of the two instruments mandated by the GMS contract: the IPQ and GPAQ. Interestingly, a recent report by the Picker Institute provides an extensive review of the quality of patient feedback questionnaires including the GPAQ but, crucially, excludes the IPQ.¹¹

M Hankins, BSc, FRSS, senior research fellow; A Fraser, BM, MRCP, FRCOG, GP registrar; A Hodson, BSc, MRCP, MRCP, GP registrar; C Hooley, MRCP, GP registrar; H Smith, DM, MRCP, FFPHM, chair of primary care, Brighton and Hove Medical School, Brighton.

Address for correspondence

Matthew Hankins, Brighton and Sussex Medical School, Mayfield House, University of Brighton, Brighton, BN1 9PH. E-mail: m.c.hankins@bsms.ac.uk

Submitted: 27 November 2006; **Editor's response:** 1 March 2007; **final acceptance:** 24 April 2007.

©British Journal of General Practice 2007; 57: 737–740.

METHOD

The MEDLINE®, AMED, CINAHL and PsychINFO databases were searched by questionnaire name as follows: 'general practice assessment questionnaire' OR 'GPAQ' OR 'general practice assessment survey' OR 'GPAS' OR 'IPQ' OR 'improving practice\$ questionnaire'. References provided by or cited on the associated web pages were also obtained. A further hand-search was carried out of papers cited in the reference section of papers obtained by either method, or papers citing those papers obtained. Only peer-reviewed papers containing empirical content relating to reliability and validity of the IPQ or GPAQ were included in the study. The search was conducted in July 2006.

RESULTS

One paper was found that met criteria for the IPQ,² and none for the GPAQ. Three papers, however, met criteria for the General Practice Assessment Survey (GPAS) questionnaire,¹²⁻¹⁴ of which the GPAQ is a shortened form.³ Results will therefore be presented for the IPQ and GPAS.

Validity and reliability of the IPQ

The version of the IPQ in the single validation paper obtained measured satisfaction with 27 statements about the patient's experience, with five response choices per item: poor, fair, good, very good, and excellent. The summary section reports that 'the IPQ has sound validity and reliability properties', but no analysis of reliability was reported. The authors investigated the internal structure of the IPQ using principal components analysis, and found two factors. This result suggests that the IPQ comprised two sets of correlated items, but it says nothing about the reliability of those sets of items.

Validity was assessed using two methods, neither of which allowed the conclusion that the validity of the IPQ was 'sound'. First, the summed scores for items 1 to 26 were found to correlate significantly with item 27 ($r = 0.78$). This questions the validity of item 27. The correlation suggests only that item 27 was likely to be measuring the same thing as the other items: it does not determine what the items were measuring.

The second method was to compare the mean satisfaction scores of younger (aged less than 40 years) and older (aged 40 years or over) patients. A significant difference was interpreted by the authors as evidence of validity, but again, this begs the question of whether the older patients were more satisfied. Social and demographic variables have been shown to influence patient satisfaction, but these effects are inconsistent between studies.¹⁵ The difference observed in this case could equally well confirm a bias in scoring due to age, which would threaten, rather than support, validity.

In summary, it was found that the claim that the reliability and validity of the IPQ were 'sound' was not supported by the data presented.

Reliability and validity of the GPAQ

The GPAQ is essentially a shortened version of the GPAS. In general, it could be expected that a shortened version of a questionnaire would retain the characteristics of the original questionnaire, and the data presented here for the GPAS are therefore approximate indicators of the reliability and validity of the GPAQ. However, ideally changes to a validated questionnaire should be followed by a pilot study assessing the effects of the changes on the performance of the questionnaire.

The GPAQ comes in two forms: one for completion with the practice, and one for postal surveys. The latter is most similar to the GPAS, which was originally designed for postal surveys. The differences between the two versions are small and largely confined to the opening section describing the purpose of the questionnaire.

In the three papers obtained, the GPAS measured patient satisfaction in several domains (for example, access and technical care). The number of items and type of response category varied from domain to domain, but all responses were coded so that higher scores indicated greater satisfaction.

Estimates of reliability for domain scores were found, reported as test-retest correlations in the range $r = 0.81$ to 0.92 and Cronbach's α coefficients in the range 0.69 to 0.95 . These estimates would typically be considered

How this fits in

Measurement of patient satisfaction is not easy and can give misleading results if questionnaires are not validated. The UK Quality and Outcomes Framework (QOF) mandates measuring patient satisfaction with either the Improving Practices Questionnaire (IPQ) or the General Practice Assessment Questionnaire (GPAQ). This review found that neither instrument has been assessed for reliability or adequately validated against external criteria. If the QOF is to have an impact on patient experience then future instruments must undergo rigorous development and testing.

acceptable. The degree of reliability of data produced by the GPAQ itself remains unknown, but could reasonably be expected to vary over the same range.

Validity was assessed in one paper as the degree of intercorrelations between scales,¹² which is not a measure of validity as no scale was known to be valid. Another study found that satisfaction with 'patient centredness', measured by a modified version of the GPAS, correlated with 'enablement' ($r = 0.51$),¹³ which is an external criterion. However the measurement of 'enablement' was not described in any detail in the paper, and 'patient centredness' was measured by collapsing scores from three of the subscales of the GPAS. It is not possible to assess the relevance of this finding to the current version of the GPAQ, which does not collapse scores in this way, when the validity of the criterion is unknown. Finally, another study found significant associations between satisfaction scores and age, socioeconomic status, and ethnicity.¹⁴ The extent to which these associations validated the GPAS, rather than demonstrated confounding effects, is unclear since their interpretation requires independent confirmation by another measure that satisfaction did indeed vary consistently with those demographic factors.

In summary, the GPAS was found to be reliable but evidence for its validity was weak. The extent to which these results generalise to the GPAQ remains unknown.

DISCUSSION

It was surprising that the two questionnaires mandated for use in the GMS contract have very little published data to support their validity, given their widespread use. No adequate assessment of reliability for either the IPQ or the GPAQ was found, and neither have been validated against an external criterion. The authors of the GPAS point out that such validation is required, but no evidence could be found that this occurred. It is important to note that this does not imply that the measures are neither reliable nor valid; but without such validation it is not clear that the questionnaires measure satisfaction at all.

Both questionnaires have been revised since their specification in the GMS contract and since the publication of the data reviewed here. Such changes would be expected to be evaluated by further validation, especially as in one case (the GPAQ) the change was made because of the apparent problematic performance of one item. Questionnaire design is an iterative process, as problems are identified and the measures improved; but the process of revision and

revalidation should, as far as possible, be exposed to peer review.

These criticisms should of course be viewed in context. If the use of surveys within the QOF was to establish the acceptability to patients of such surveys, then issues of reliability and validity are of less importance at this stage. If the intention was to give a rough guide to the level of patient satisfaction with general practice, then these criticisms apply more to the failure in the process of questionnaire validation than to the data generated. The robustness of the measure of patient satisfaction data would appear to contrast poorly with the sound evidence base of QOF indicators that relate to disease management. However, it could be argued that there is virtue in engaging the primary care team in considering the patient's experience of care, and patient satisfaction surveys can act as the catalyst. Whether this alternative agenda warrants the time and resources put into surveys or is the most appropriate way to raise the profile of patient satisfaction is a matter for debate. If the data are to be used for comparison of practices, or of practitioners within practices, or to demonstrate improvements in patient satisfaction over time, then the validity and precision of measurement are hugely important. This applies equally to the use of patient survey data in any GP education, appraisal, and revalidation exercises.

It is possible that this study has failed to obtain all the available validation data for these questionnaires. It focused on publicly-available information and restricted the search to peer-reviewed articles, thus excluding the 'grey literature' — any internal reports. Access to IPQ and GPAQ datasets may have helped to evaluate reliability, assuming that the data were collected using sound and standardised sampling techniques.

There are currently proposals to simplify the patient survey component of the QOF.¹⁶ It is recommended that, whatever the final form of future patient surveys, extensive piloting should take place to ensure that the validity and reliability are 'fit for purpose'. The almost universal engagement of general practices in the patient survey process provides a valuable opportunity to improve and refine survey materials and help the QOF have an impact on patient experience.

Funding body

Not applicable

Ethics committee

Not applicable

Competing interests

The authors have stated that there are none

Acknowledgements

The time Alice Fraser, Andrew Hodson and Claire Hooley spent in the academic department was organised by Professor Abdol Tavabie and funded by the Kent, Surrey and Sussex Deanery as part of a GP Registrar Extension Scheme.

REFERENCES

1. British Medical Association/The NHS Confederation. *The new GMS contract*. London: British Medical Association and NHS Confederation, 2003.
2. Greco M, Powell R, Sweeney K. The Improving Practice Questionnaire (IPQ): a practical tool for general practices seeking patient views. *Educ Prim Care* 2003; **14**: 440–448.
3. National Primary Care Research and Development Centre. *General Practice Assessment Questionnaire (GPAQ)*. Manchester: National Primary Care Research and Development Centre, Manchester University.
4. Ross CK, Steward CA, Sinacore JM. A comparative study of seven measures of patient satisfaction. *Med Care* 1995; **33**: 392–406.
5. Hays RD, Ware JE. My medical care is better than yours: social desirability and patient satisfaction ratings. *Med Care* 1986; **24**: 519–525.
6. Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use*. 3rd edn. Oxford: Oxford University Press, 2003.
7. Bowling A, Ebrahim S. *Handbook of health research methods*. Buckingham: Open University Press, 2005.
8. Carmines EG, Zeller RA. *Reliability and validity assessment*. Newbury Park: Sage Publications, 1991.
9. Guilford JP. *Psychometric methods*. New York: McGraw-Hill, 1954.
10. Sitzia J. How valid and reliable are patient satisfaction data? An analysis of 195 studies. *Int J Qual Health Care* 1999; **11**: 319–328.
11. Chisholm A, Askham J. *What do you think of your doctor?* Oxford: Picker Institute Europe, 2006.
12. Ramsay J, Campbell JL, Schroter S, Green J, Roland M. The general practice assessment survey (GPAS): Tests of data quality and measurement properties. *Fam Pract* 2000; **17**: 372–379.
13. Bower P, Mead N, Roland M. What dimensions underlie patient responses to the general practice assessment survey? A factor analytic study. *Fam Pract* 2002; **19**: 489–495.
14. Bower P, Roland MO. Bias in patient assessments of general practice: General Practice Assessment Survey scores in surgery and postal responders. *Br J Gen Pract* 2003; **53**: 126–128.
15. Carr-Hill RA. The measurement of patient satisfaction. *J Public Health Med* 1992; **14**: 236–249.
16. British Medical Association/The NHS Confederation. *Revisions to the GMS contract, 2006/07: delivering investment in general practice*. London: British Medical Association/The NHS Confederation.