

Patients' experience and satisfaction in primary care: secondary analysis using multilevel modelling

Chris Salisbury, professor of primary health care,¹ Marc Wallace, ST2 core medical doctor,² Alan A Montgomery, reader in health services research¹

¹University of Bristol, Bristol BS8 2AA, UK

²Royal Berkshire Hospital, Reading RG1 5AN, UK

Correspondence to: C Salisbury
c.salisbury@bristol.ac.uk

Cite this as: *BMJ* 2010;341:c5004
doi:10.1136/bmj.c5004

ABSTRACT

Objective To explore whether responses to questions in surveys of patients that purport to assess the performance of general practices or doctors reflect differences between practices, doctors, or the patients themselves.

Design Secondary analysis of data from a study of access to general practice, combining data from a survey of patients with information about practice organisation and doctors consulted, and using multilevel modelling at practice, doctor, and patient level.

Setting Nine primary care trusts in England.

Participants 4573 patients who consulted 150 different doctors in 27 practices.

Main outcome measures Overall satisfaction; experience of wait for an appointment; reported access to care; satisfaction with communication skills.

Results The experience based measure of wait for an appointment was more discriminating between practices (practice level accounted for 20.2% (95% confidence interval 9.1% to 31.3%) of variance) than was the overall satisfaction measure (practice level accounted for 4.6% (1.6% to 7.6%) of variance). Only 6.3% (3.8% to 8.9%) of the variance in the doctors' communication skills measure was due to differences between doctors; 92.4% (88.5% to 96.4%) of the variance occurred at the level of the patient (including differences between patients' perceptions and random variation). At least 79% of the variance on all measures occurred at the level of the patient, and patients' age, sex, ethnicity, and housing and employment status explained some of this variation. However, adjustment for patients' characteristics made very little difference to practices' scores or the ranking of individual practices.

Conclusions Analyses of surveys of patients should take account of the hierarchical nature of the data by using multilevel models. Measures related to patients' experience discriminate more effectively between practices than do measures of general satisfaction. Surveys of patients' satisfaction fail to distinguish effectively between individual doctors because most of the variation in doctors' reported performance is due to differences between patients and random error rather than differences between doctors. Although patients' reports of satisfaction and experience are systematically

related to patients' characteristics such as age and sex, the effect of adjusting practices' scores for the characteristics of their patients is small.

INTRODUCTION

Surveys of patients are increasingly used internationally as an indicator of the performance of health systems. In some countries, the results of surveys are used within "pay for performance" schemes.^{1,2} Under the national quality and outcomes framework, general practices in the United Kingdom receive some of their income according to the results of a national survey of patients' experiences.³

Being sure that patients' responses in surveys act as a reliable indicator of performance is therefore important. Patients can describe high levels of satisfaction at the same time as describing experiences that are suboptimal,⁴ and patients' subjective satisfaction varies systematically with certain characteristics such as the age, sex, and ethnicity of the patient.⁵⁻⁷ Whether this is because of differences in expectation, differences in the service provided to patients with different characteristics, or differences in the way patients report their experiences is unclear.⁸

In response to these problems, a recent trend has favoured questions about patients' experiences (box 1).⁴ This is based on the assumption that reported experience should be less influenced by subjective expectation than is reported satisfaction. However, whether reports of patients' experiences are also systematically associated with sociodemographic characteristics is not clear.⁹

An ideal measure of patients' experiences for use in general practice should show variation between high performing and low performing practices, with less variation between patients within practices than would be anticipated for questions about satisfaction. Conversely, if a measure shows little variation by practice this would imply that the measure is an unreliable indicator of practices' performance.

Many surveys of patients include questions about individual doctors as well as questions about practices' organisation. In theory, if these questions are reliable, scores obtained from questions about doctors' performance should show considerable variation between

Box 1 Examples of satisfaction questions and experience questions**Experience questions**

These reflect actual experience, aiming to avoid value judgments and the effects of existing expectations. For example:

- Were you able to get an appointment within two working days?
- How long after your appointment time do you normally wait to be seen?

Satisfaction questions

These are subjective and often non-specific. For example:

- How satisfied are you with the appointment system in your practice?
- How do you rate your doctor's caring and concern for you?

different doctors, with “good” doctors getting consistently high scores from patients and “poor” doctors getting consistently low scores. However, patients' reports about doctors may reflect their satisfaction with their practice, or vice versa, as well as reflecting the patient's own characteristics. In other words, if patients think their practice is good they may tend to say that the doctor is good, in a form of the “halo effect.”¹⁰ Alternatively, if patients are dissatisfied with aspects of the practice's performance they may express less satisfaction with their doctor. These considerations about whether variation in patients' responses reflect practices' or doctors' performance are important, because findings from satisfaction surveys are increasingly used in the appraisal and revalidation of individual doctors.¹¹

Surveys of patients who have recently attended an appointment generate data that have a hierarchical or multilevel structure. Patients are “nested” within doctors, who are “nested” within practices. Analysis of this type of data should therefore use multilevel modelling approaches that take appropriate account of this clustered nature of the data and enable exploration of the sources of variation at each level.¹² In the context of this study, this is important because different doctors and practices are likely to attract patients with particular characteristics.

The aim of this study was to use a multilevel modelling approach to simultaneously explore the extent to which factors at the level of the practice, the doctor, and the patient determine measures of patients' satisfaction and experience that purport to reflect the performance of individual doctors or practices. The key hypotheses were that more variation in practices' outcomes would occur at practice level than at doctor level, that more variation in outcomes intended to measure the performance of individual doctors would occur at doctor level than at practice level, and that less variation would exist at the level of the patient in responses to questions based on patients' experience than for patients' satisfaction.

In addition, we used multilevel modelling approaches to identify the most important variables that explain variation at each level. Although a considerable literature exists about patients' sociodemographic factors associated with satisfaction in primary care,⁵⁻⁷ and also about characteristics of general

practitioners or practices associated with patients' satisfaction,^{13,14} few previous studies have appropriately taken account of the three level hierarchical structure in which patients are clustered within doctors and then within general practices.^{15,16} Multilevel models are increasingly being used to examine sources of variation at different levels of organisation in the health service—for example, how consumer assessment scores vary across health plans, medical groups, and providers.¹⁶⁻²⁰

Finally, we have used the findings to explore the effect on practices' scores of adjusting for patients' characteristics. This form of adjustment is sometimes used in surveys of hospital patients and has been debated with regard to surveys of general practice patients.^{8,21,22}

METHODS

This paper is based on a secondary analysis of data obtained from a survey of patients' satisfaction done in 47 practices that took part in an evaluation of the advanced access initiative in 2005-6. Full details of the survey have been published elsewhere.²³ Briefly, patients who consulted each practice over several consecutive days completed a survey after they had seen a doctor, with a target of at least 100 completed questionnaires from each practice. Non-responders were followed up by a postal reminder. The questionnaire included the general practice assessment questionnaire instrument (version 1),²⁴ one of the validated questionnaires approved at the time to justify payments to practices under the quality and outcomes framework.³ The overall response rate for the survey was 84%.²³

Within the evaluation, we collected data about a range of variables relating to the structure and organisation of each practice.²⁵ These variables included the number of patients registered on the practice list, the number of full time equivalent doctors, whether the practice was approved for postgraduate training, whether the practice operated an “advanced access” appointment system, and other variables as listed in box 2.

Information about individual doctors came from the General Medical Council's online register of medical practitioners (www.gmc-uk.org/), supplemented where necessary by direct contact with the relevant practices. Variables included the doctors' sex, year of qualification, and whether their first medical qualification was obtained in the UK. Information about patients' characteristics came from the questionnaire, and box 2 also shows details of these variables.

The advanced access evaluation did not require practices to record the identity of the doctor being consulted by the patient completing the questionnaire. This paper is based on the 27 practices that did collect data about the individual doctors, making it possible to do multilevel analysis incorporating variables at all three levels of patient, doctor, and practice.

Question items in the general practice assessment questionnaire can be analysed individually, and some question items can be combined to generate several

Box 2 Explanatory variables included in multilevel models**Practice**

- List size
- Number of full time equivalent general practitioners
- Operating advanced access
- Ex-fundholding practice
- Total points under quality and outcomes framework
- Operating under Personal Medical Services contract
- Training practice
- Receives dispensing payments

Doctor

- Sex
- Years since qualification
- Whether qualified in UK

Patient

- Age
- Sex
- Ethnicity
- Housing status
- Employment status

scales.²⁴ We selected, a priori, two questions and two scales to test our hypotheses, as follows. We chose a single question about overall satisfaction with the practice to represent global subjective patients' satisfaction. We chose a single question about how long it usually takes to get an appointment with any doctor to represent a question about patients' experience. We chose the questionnaire's "access" scale to assess patients' assessment of a particular aspect of a practice's performance. It comprises questions in which patients are asked to rate their experience in relation to specific dimensions of access (for example, how quickly they can usually see any doctor). Although the ratings involve value judgments, like conventional satisfaction questions, the designers of the instrument have linked these evaluations to specific experiences. We chose the "communication" scale to represent patients' satisfaction with the communication skills of the particular doctor they had consulted that day. Table 1 shows further details of these outcomes and how they were scored. Responses on the two single questions were scored on ordinal scales. As the distributions of responses on these questions were skewed, we did a

normal score transformation for these two questions before analysis by using the "NSCO" command in MLwiN to meet the assumption of normality on which regression models are based. This transformation assigns expected values from the standard normal distribution according to the ranks of the original scores. For all four outcomes, the responses were scored so that higher scores represent greater satisfaction.

Analysis

We chose practice, doctor, and patient related variables for initial inclusion on the basis of a theoretical justification or evidence from previous literature indicating that they might relate to patients' satisfaction or experience. We then did regression analysis using backward selection to examine associations between these variables and each of the four outcomes, which we labelled "overall satisfaction," "wait for appointment," "access," and "communication," eliminating variables with P values above 0.1. We retained all variables in the models for at least one of the outcomes, so all went forward to the multilevel analysis. We treated data hierarchically as follows: patients (level 1), doctors (level 2), and practices (level 3). We used Stata version 11 and MLwiN version 2.20 for all analyses.

As one aim was to compare models that were or were not adjusted for explanatory variables, we needed to ensure that we included the same patients in the different models for each outcome. We therefore included patients in the final multilevel models only if they had complete data on all the explanatory variables. However, the number of patients providing data for each satisfaction/experience outcome varied, so comparisons between the different outcomes should be made with caution.

We constructed multilevel models for each outcome, treating practice and doctor levels as random effects. We then added explanatory variables at practice, doctor, and patient levels as fixed effects. We did multilevel analysis on each outcome by using the iterative generalised least squares method of estimation. We calculated variance partition coefficients for each outcome. Variance partition coefficients represent the proportion of total variance in an outcome that is due to differences occurring at each level. In other words, a high coefficient at practice level indicates that more of the variation in the model is due to differences between practices than between doctors or between patients.

Table 1 | Questions and scales used as outcomes in analysis

Variable	Question wording	Scoring
Overall satisfaction	All things considered, how satisfied are you with your practice?	Seven point scale from "completely satisfied" to "completely dissatisfied"
Wait for appointment	Thinking of times when you are willing to see any doctor: how quickly do you usually get seen?	Six point scale from "same day" to "five or more working days," with additional "does not apply" option.
Access	Scale created from six separate questions ²⁴ relating to contacting the practice or making an appointment	Scale from 0 to 100
Communication	Scale created from eight questions ²⁴ about patient's satisfaction with communication with the doctor in today's consultation	Scale from 0 to 100

We calculated variance partition coefficients first in a model with random intercepts at the practice and doctor levels and no explanatory variables (raw coefficients) and then after adjustment for patients' characteristics (patient adjusted coefficients), and finally after adjustment for characteristics of the practice, doctor, and patient (fully adjusted coefficients). The coefficients from the adjusted models show the relation between each explanatory variable and the outcome, and the differences between the raw and adjusted models show the extent to which the explanatory variables explain variation in the outcome. We compared the sets of models for each outcome by using likelihood ratio tests. We also compared the age and sex of patients who responded and did not respond to the questionnaire and between those included in the final dataset and those not included (because of non-response or missing data).

RESULTS

Of 7195 eligible patients, 6045 (84.0%) responded, of whom 5496 (90.9%) had an identifier to signify the general practitioner. In 5150 cases (93.7% of those with an identifier; 85.2% of all respondents) we were able to identify the general practitioner in the Medical Directory. These 5150 patients provided the dataset for analysis. We included a total of 16 potential explanatory variables—eight at the level of the practice, three at the level of the doctor, and five at the level of the patient (box 2).

After exclusion of patients with missing data on any of these variables, the final dataset included 27 practices, 150 doctors, and 4573 patients (a mean of 169 (range 52-323; SD 80.1) patients per practice and 30 (1-104; 20.2) patients per doctor). No differences existed between respondents and non-respondents or between those patients finally included or excluded, in terms of age and sex. The web appendix shows the characteristics of the sample in terms of the explanatory variables.

Table 2 shows descriptive statistics for the range of practices' scores on each outcome. This simple, single level analysis shows the extent of variation in practices' performance on these measures.

Figure 1 illustrates this variation between practices, showing practices' residuals and 95% confidence limits for "overall satisfaction" and "wait for appointment" (an "experience" measure). A residual is the difference between an observed score and the score predicted by a regression equation, so plots such as these show how much the individual units (in this case practices) differ from the mean. For the overall satisfaction measure, only nine of the 28 practices had 95% confidence limits that excluded the average, whereas the experience based wait for appointment measure was much more discriminating between practices.

Table 3 shows the variance components models. The variance partition coefficients show the proportion of variance occurring at the level of practice, doctor, or patient for each of the outcomes in the unadjusted and adjusted models. This table shows,

Table 2 | Descriptive statistics for practices' scores on four different measures of satisfaction

Outcome	Practice scores* (n=27)	
	Mean (SD)	Range
Overall satisfaction:		
As recorded (scored 1-8)	6.00 (0.23)	5.60-6.44
Normalised	0.01 (0.21)	-0.33-0.40
Wait for appointment:		
As recorded (scored 1-7)	4.94 (0.60)	3.75-5.89
Normalised	-0.09 (0.38)	-0.73-0.54
Access (scored 0-100)	63.4 (7.46)	48.3-80.6
Communication (scored 0-100)	83.0 (3.51)	76.8-89.4

*Mean and range of 27 practices' scores, which are themselves means of patients' scores per practice.

for example, that 20.2% of the variance in the wait for appointment outcome was due to differences between practices, 0.8% to differences between the doctors consulted, and the remaining 79.1% to variance at the level of the patient, which includes differences between individual patients' perceptions and random error. After inclusion of information about patients' characteristics, the total amount of unexplained variation decreased very slightly (see variance denominator) but the proportion due to variation at the level of the practice, doctor, and patient did not change. However, further adjustment for practice and doctor related variables substantially reduced the proportion of variance due to differences between practices (because some of the variation has now been explained by known factors such as the list size of the practice).

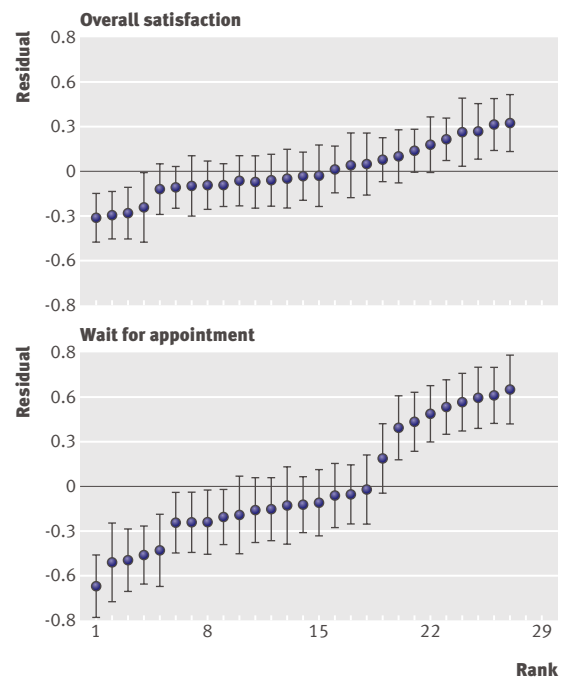


Fig 1 | Plot of practices' residuals (difference between observed score and score predicted by regression equation), with 95% confidence intervals, for overall satisfaction (left) and wait for appointment (right)

Table 3 | Variance at level of practice, doctor, and patients for each outcome, before and after adjustment for explanatory variables

Outcome	Total variance	Estimated variance (95% CI) as percentage of total variance in outcome		
		Between practice	Between doctor	Patient plus random
Overall satisfaction (n=4414):				
Unadjusted	0.719	4.6 (1.6 to 7.6)	1.5 (0.4 to 2.6)	93.9 (89.8 to 98.0)
Adjusted for patients' characteristics	0.701	4.9 (1.8 to 7.9)	1.0 (0 to 2.1)*	94.2 (90.2 to 98.1)
Fully adjusted for practice, doctor, and patient related variables	0.686	3.4 (1.1 to 5.6)	0.4 (-0.4 to 1.3)*	96.2 (92.2 to 100.2)
Wait for appointment (n=4058):				
Unadjusted	0.654	20.2 (9.1 to 31.3)	0.8 (0 to 1.7)*	79.1 (75.5 to 82.6)
Adjusted for patients' characteristics	0.650	20.2 (9.0 to 31.3)	0.8 (0 to 1.7)*	79.1 (75.5 to 82.7)
Fully adjusted for practice, doctor, and patient related variables	0.560	7.3 (2.8 to 11.9)	0.9 (0 to 1.9)*	91.8 (87.6 to 96.0)
Access (n=4517):				
Unadjusted	325.9	14.9 (6.4 to 23.3)	1.5 (0.4 to 2.6)	83.7 (80.2 to 87.2)
Adjusted for patients' characteristics	321.0	15.5 (6.7 to 24.3)	1.4 (0.3 to 2.4)	83.1 (79.7 to 86.6)
Fully adjusted for practice, doctor, and patient related variables	299.7	9.7 (4.0 to 15.3)	1.3 (0.2 to 2.3)	89.1 (85.4 to 92.8)
Communication (n=4423):				
Unadjusted	306.5	1.2 (0 to 3.0)†	6.3 (3.8 to 8.9)	92.4 (88.5 to 96.4)
Adjusted for patients' characteristics	298.5	1.4 (0 to 3.2)†	5.5 (3.1 to 7.8)	93.1 (89.2 to 97.1)
Fully adjusted for practice, doctor, and patient related variables	294.7	1.6 (0 to 3.3)†	4.0 (2.1 to 6.0)	94.3 (90.3 to 98.3)

Likelihood ratio test: all adjusted models $P < 0.001$ compared with null model.

*Lower confidence limit restricted to 0.

With regard to our first hypothesis (that more variation in practice outcomes would occur at practice level than at doctor level), table 3 shows that for overall satisfaction and communication a very small proportion of the variation occurs at practice level, but for access and wait for appointment a larger proportion is attributable to practices. The low variance partition coefficients at the level of the doctor for overall satisfaction, access, and wait for appointment indicate that patients' assessments of a practice's performance were not influenced by which individual doctor they consulted.

With regard to our second hypothesis (that more variation in outcomes intended to measure the performance of individual doctors would occur at doctor level than at practice level), we found evidence that more of the variation in results on the communication scale, which was intended to assess individual doctors,

occurred at the level of the doctor rather than the practice. However, only 6% of variation (unadjusted model) occurred at the level of the doctor, and most of the variation occurred at the level of the patient. This includes variation between the reports of individual patients as well as measurement error and unexplained random variation. Figure 2 shows that the residuals for individual doctors almost all overlapped the mean.

With regard to our third hypothesis (that less variation would exist at the level of the patient in responses to questions based on patients' experience than for patients' satisfaction), we found evidence that scores on measures relating to specific experiences of patients (wait for appointment, access scale) had less variation at patient level and more variation at practice level than did scores on measures of patients' satisfaction (overall satisfaction, communication). For all models, most of the variation occurred at the level of the patient and remained unexplained even after adjustment for all of the practice, doctor, and patient related explanatory variables included.

Table 4 provides details of relations between the explanatory variables at the level of the practice, doctor, and patient and the four outcomes. In combination with table 3, this shows that several factors related to the organisation of the practice had a significant association with practices' performance in relation to the ability to get an appointment but not the other outcomes. For doctors, being qualified in the UK and being qualified for fewer years was associated with higher patients' scores for communication. The influence of patients' characteristics varied for each outcome, but patients' age was associated with three of

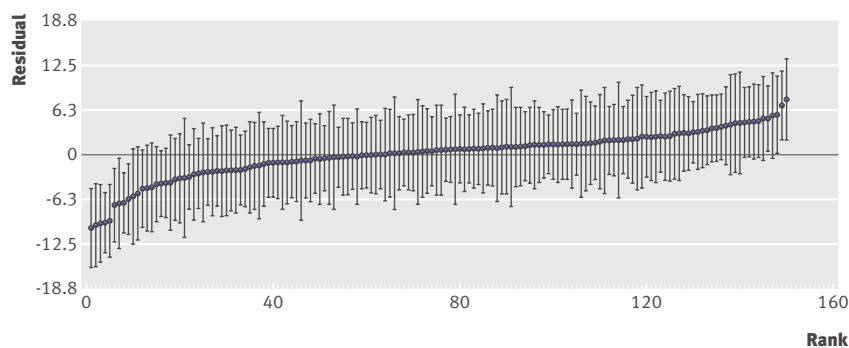


Fig 2 | Plot of residuals (difference between observed score and score predicted by regression equation) at level of doctor for communication, with 95% confidence intervals

Table 4 | Relation between potential explanatory variables and each outcome in multilevel models

Variables	Overall satisfaction— patient level mean 0.00 (SD 0.85)		Ability to get an appointment—patient level mean -0.05 (SD 0.81)		Access—patient level mean 63.10 (SD 17.75)		Communication—patient level mean 83.10 (SD 17.50)	
	Coefficient (95% CI)	P value	Coefficient (95% CI)	P value	Coefficient (95% CI)	P value	Coefficient (95% CI)	P value
Practice level								
List size (per 1000 patients)	0.01 (0.00 to 0.01)	0.32	0.13 (0.05 to 0.21)	0.001	0.82 (-1.23 to 2.86)	0.44	0.68 (-0.49 to 1.85)	0.25
No of full time equivalent GPs	-0.07 (-0.19 to 0.05)	0.24	-0.21 (-0.36 to -0.06)	0.01	-3.43 (-7.33 to 0.47)	0.08	-1.55 (-3.8 to 0.70)	0.18
Advanced access practice*	0.06 (-0.09 to 0.21)	0.41	0.31 (0.12 to 0.50)	<0.001	3.95 (-1.01 to 8.92)	0.12	0.70 (-2.04 to 3.45)	0.62
Ex-fundholding practice*	0.02 (-0.12 to 0.16)	0.79	0.17 (-0.01 to 0.34)	0.06	0.40 (-4.19 to 4.99)	0.86	-1.14 (-3.64 to 1.35)	0.37
QOF points (log)†	0.02 (-0.15 to 0.20)	0.80	0.25 (0.02 to 0.48)	0.03	1.91 (-3.94 to 7.76)	0.52	3.43 (0.13 to 6.74)	0.04
PMS practice*	-0.11 (-0.28 to 0.05)	0.19	0.11 (-0.11 to 0.32)	0.33	-3.43 (-8.97 to 2.11)	0.23	1.30 (-1.79 to 4.39)	0.41
Training practice*	0.14 (-0.01 to 0.29)	0.07	-0.02 (-0.22 to 0.18)	0.85	2.83 (-2.30 to 7.96)	0.28	-0.38 (-3.23 to 2.47)	0.79
Practice receives dispensing payments*	0.13 (-0.06 to 0.32)	0.18	-0.12 (-0.37 to 0.12)	0.33	4.10 (-2.42 to 10.61)	0.22	0.78 (-2.75 to 4.30)	0.67
Doctor level								
Male sex*	0.07 (0.00 to 0.13)	0.04	0.03 (-0.03 to 0.09)	0.32	1.25 (-0.17 to 2.67)	0.08	1.38 (-0.40 to 3.16)	0.13
Years since qualification	0.00 (0.00 to 0.01)	0.32	0.00 (-0.01 to 0.00)	0.32	0.00 (-0.08 to 0.08)	1.00	-0.11 (-0.21 to -0.01)	0.04
UK qualified*	0.08 (-0.01 to 0.17)	0.07	-0.04 (-0.12 to 0.05)	0.41	-0.17 (-2.20 to 1.86)	0.87	4.36 (1.81 to 6.90)	<0.001
Patient level								
Age (years)	0.01 (0.00 to 0.01)	<0.001	0.00 (0.00 to 0.00)	1.00	0.09 (0.06 to 0.12)	<0.001	0.11 (0.07 to 0.14)	<0.001
Male sex*	-0.02 (-0.07 to 0.03)	0.44	-0.10 (-0.14 to -0.05)	<0.001	1.25 (0.25 to 2.25)	0.01	0.39 (-0.64 to 1.43)	0.46
White ethnicity*	0.13 (-0.03 to 0.29)	0.12	0.13 (-0.02 to 0.28)	0.09	-0.07 (-3.29 to 3.15)	0.96	4.56 (1.25 to 7.87)	0.01
Housing status: owner/occupier*	0.01 (-0.04 to 0.07)	0.68	0.03 (-0.02 to 0.08)	0.20	-1.20 (-2.32 to -0.08)	0.03	1.68 (0.53 to 2.84)	<0.001
Employed*	-0.05 (-0.10 to 0.01)	0.10	-0.06 (-0.11 to 0.00)	0.03	-1.91 (-3.04 to -0.77)	<0.001	-0.24 (-1.41 to 0.93)	0.69

For each outcome measure, higher scores represent greater satisfaction.

GP=general practitioner; PMS=Personal Medical Services; QOF=quality and outcomes framework.

*Binary variable: yes=1, no=0.

†QOF scores were log transformed because of skewed distribution.

the four outcomes, and sex, ethnicity, housing status, and employment status were each associated with at least one outcome.

Although we found evidence that these characteristics of patients were associated with the outcomes, the coefficients were generally small relative to overall mean scores. As can be seen from table 3, adding patients' explanatory variables to the null models did not reduce the unexplained variation at the level of the patient or increase the proportion of variation explained at practice level. Figure 3 shows the practices' residuals for the access scale with and without adjustment for patients' characteristics, with practices ranked by unadjusted access score. This shows that adjusting for patients' characteristics makes very little difference to practices' scores or to the performance of individual practices relative to other practices.

DISCUSSION

This paper describes several important findings relating to the interpretation of surveys of patients' satisfaction and experience. It provides support for the ability of questions of the type used in the general practice assessment questionnaire to discriminate between aspects of the performance of practices and of individual doctors within those practices. It also provides support for the concept that questions about patients' experience provide a more discriminating measure of a practice's performance than do subjective questions about satisfaction.

The fact that 20% of variation in patients' reported experience of the wait for an appointment occurs at the level of the practice is a strong endorsement for the use of this type of question as a measure of a practice's performance. In addition, the reduction in unexplained variation at practice level once the model was adjusted for a range of practice and doctor related factors implies that much of the variation between practices can be explained by these factors (detailed in table 4), some of which may be modifiable and used to drive improvement in quality.

This study has also shown the usefulness of multi-level modelling to explore sources of variation. Multi-level modelling is widely used in education—for example, to explore the value of school league tables after taking account of the characteristics of pupils²⁶—and is very relevant to many studies in health services research, as patients' data are similarly clustered at multiple levels. By taking appropriate account of the hierarchical nature of the data, this study has provided estimates of the influence of practice, doctor, and patient related characteristics on patients' satisfaction and experience that are likely to be more realistic than those from earlier single level studies.

Limitations

The practices included in this study were not randomly selected, and their willingness to take part in a research study of appointment systems may mean that they are not necessarily representative of all practices. The

number of practices included is also relatively small. Patients were included in the study because they attended a consultation, so the findings will be weighted towards those who attend most often; those who do not attend are not represented. An alternative approach would be to post a survey to a random sample of all patients, which would include those who do not use the service. However, postal questionnaire surveys typically lead to much lower response rates, which may introduce non-response bias.

The need to exclude patients with missing data on explanatory variables meant that only 88.8% of the patients originally included in the dataset were included in analyses. This may introduce bias if data are not missing completely at random. We explored methods to impute missing data, but these are not yet well developed for models with more than two levels. The loss of data due to missing data also reduces the statistical power to detect differences, as does the skewed nature of the outcome variables.

Implications

Measures of patients' satisfaction discriminate poorly between practices or doctors, because random error and differences in people's perceptions account for more than 90% of the variance. This is consistent with earlier research using multilevel models to examine patients' satisfaction with hospital care, which showed that less than 5% of the variation in overall satisfaction occurred at the level of hospital or department,²⁷ and also with a previous study in primary care which suggested that 90-97% of variation in different satisfaction outcomes occurred at the level of the patient.¹⁷ Conversely, Haggerty found that 20% of variance in patients' reports of accessibility of general practice in Canada could be explained at the level of the practice and 3% at the level of the doctor,¹⁵ and a study of patients in California visiting primary care physicians suggested that about between 28% and 48% of variation was due to system related factors, with more variation being due to differences between doctors than to differences between medical groups or localities.¹⁶ Taken together with our findings, the studies suggest

that scores based on questions about specific aspects of organisation of care (particularly access to care) are more likely to vary between practices than between doctors, and such questions are more discriminatory than are questions about general satisfaction. Measures of doctor-patient interaction are more likely to vary between doctors than between practices, but considerable variation at the level of the patient and random variation exist. For the communication scale in our study, so little variation exists at the level of the doctor that the reliability of using this type of measure to assess an individual doctor's performance is questionable. In the vast majority of cases, meaningfully distinguishing between doctors is impossible, although plots such as that in figure 3 do allow attention to be focused on the small number of doctors at the extremes who seem to have scores considerably above or below the mean.

The finding that patients' characteristics influence responses to questions about experience of health services as well as satisfaction raises the question about whether this reflects different expectations or differences in the care provided to different types of patients within the same practices. This is an important debate.²² If patients' experience is related to expectation rather than to performance of the practice, then failing to adjust practices' scores for the characteristics of the population of patients could lead to systematic misrepresentation of the performance of practices that cater for particular patient groups, such as those from ethnic minorities.⁸ This is particularly a concern when pay is linked to scores from surveys, as practices working in challenging circumstances could be further disadvantaged by loss of investment. However, if the lower scores reported by certain types of patient reflect a lower quality of care, then adjusting practices scores would mean that inequitable care provision would not be identified.²⁸

This paper shows that although patients' characteristics influence their responses with regard to both satisfaction and experience, the overall effect on practices' scores is small, at least in the practices included in this study. A study of health plans in the Netherlands suggested that adjustment for patient related factors other than age had little impact,²⁰ although a study of satisfaction with inpatient care suggested that it did for a small number of hospitals.²¹ Our findings should therefore be replicated in a larger sample of practices, as results may be different in practices with atypical populations, particularly those with a high proportion of young or ethnic minority patients.

Conclusions

Analyses of surveys of patients should take account of the hierarchical nature of the data by using multilevel models and explore the effect of explanatory variables at the levels of the practice, doctor, and patient. Measures of patients' experience discriminate more effectively between practices than do measures of satisfaction. The high level of variability between patients and random error means that surveys of patients' satisfaction are unlikely to effectively discriminate between individual doctors. Reports of

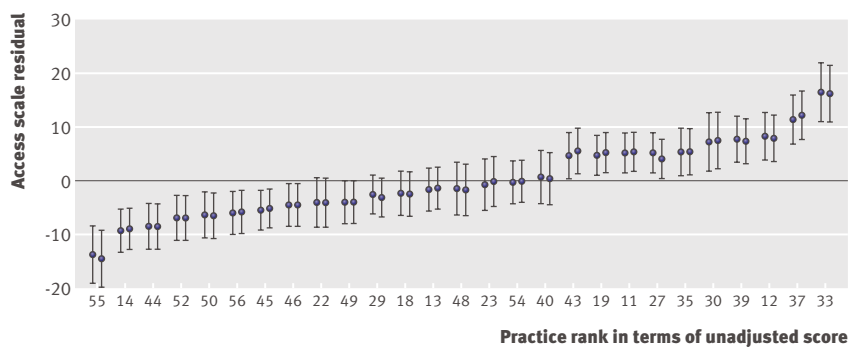


Fig 3 | Practice residuals (difference between observed score and score predicted by regression equation) on access scale before and after adjustment for patients' characteristics. Each pair of lines represents point estimate and confidence limits for a practice, before and after adjustment

WHAT IS ALREADY KNOWN ON THIS TOPIC

Surveys of patients are used to assess the performance of doctors and practices, and these increasingly enquire about patients' specific experiences as well as their satisfaction

Few studies have explored the extent to which variation in reported satisfaction and experience is due to differences between practices, doctors, or patients themselves

Few studies have quantified the effect on practices' scores of adjusting for patients' characteristics

WHAT THIS STUDY ADDS

Questions about patients' satisfaction discriminate poorly between practices and doctors, but questions about specific experiences are more discriminatory

Adjusting for patients' characteristics makes little difference to practices' performance scores

patients' experience as well as satisfaction are systematically related to patients' characteristics such as age and ethnicity, but the effect of adjusting practices' scores for the characteristics of their patients is small.

We thank the patients and practices that participated and the Advanced Access Evaluation Team that did the research on which this analysis is based.

Contributors: CS led the evaluation of advanced access, with methodological support from AAM. CS had the idea for this paper. CS and MW did the analysis, with advice from AAM. CS wrote the paper, with comments from MW and AAM. All authors revised and approved the final paper. CS is the guarantor.

Funding: The evaluation of advanced access was funded by the NHS Research and Development Programme on Service and Delivery Organisation (ref SDO/70/2004). The views expressed in this publication are those of the authors and not necessarily those of the funders. This secondary analysis had no specific funding. The funders had no input into study design; the collection, analysis, or interpretation of data; the writing of the report; or the decision to submit the article for publication. The authors are independent from the research funders.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that CS, MW, and AM, and their spouses, partners, and children, have no financial or non-financial interests that may be relevant to the submitted work.

Ethical approval: Thames Valley Multicentre Research Ethics Committee approved this study (ref 04/12/024).

Data sharing: No additional data available.

- Roland M, Elliott M, Lyraztopoulos G, Barbieri J, Parker RA, Smith P, et al. Reliability of patient responses in pay for performance schemes: analysis of national General Practitioner Patient Survey data in England. *BMJ* 2009;339:b3851.
- Salisbury C. Using patient experience within pay for performance programmes. *BMJ* 2009;339:b4224.
- NHS Employers. Quality and Outcomes Framework guidance for GMS contract 2009/10: delivering investment in general practice. NHS Employers, 2009 (available at www.nhsemployers.org/Aboutus/Publications/Documents/QOF_Guidance_2009_final.pdf).
- Jenkinson C, Coulter A, Bruster S, Richards N, Chandola T. Patients' experiences and satisfaction with health care: results of a questionnaire study of specific aspects of care. *Qual Saf Health Care* 2002;11:335-9.
- Hall JA, Dorman MC. Patient socio-demographic characteristics as predictors of satisfaction with medical care: a meta-analysis. *Soc Sci Med* 1990;30:811-8.
- Williams SJ, Calnan M. Key determinants of consumer satisfaction with general practice. *Fam Pract* 1991;8:237-42.
- Campbell JL, Ramsay J, Green J. Age, gender, socioeconomic, and ethnic differences in patients' assessments of primary health care. *Qual Health Care* 2001;10:90-5.
- Mead N, Roland M. Understanding why some ethnic minority patients evaluate medical care more negatively than white patients: a cross sectional analysis of a routine patient survey in English general practices. *BMJ* 2009;339:b3450.
- Bleich SN, Ozaltin E, Murray CK. How does satisfaction with the health-care system relate to patient experience? *Bull World Health Organ* 2009;87:271-8.
- Thorndike EL. A constant error in psychological ratings. *J Appl Psychol* 1920;4:25-9.
- Department of Health. Appraisal for GPs. 2007. <http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Managingyourorganisation/Workforce/EducationTrainingandDevelopment/Appraisals/index.htm>.
- Goldstein H. Multilevel statistical models. 3rd ed. Arnold, 2009.
- Sans-Corrales M, Pujol-Ribera E, Gene-Badia J, Pasarín-Rua MI, Iglesias-Perez B, Casajuana-Brunet J. Family medicine attributes related to satisfaction, health and costs. *Fam Pract* 2006;23:308-16.
- Campbell SM, Hann M, Hacker J, Burns C, Oliver D, Thapar A, et al. Identifying predictors of high quality care in English general practice: observational study. *BMJ* 2001;323:784-7.
- Haggerty JL, Pineault R, Beaulieu MD, Brunelle Y, Gauthier J, Goulet F, et al. Practice features associated with patient-reported accessibility, continuity, and coordination of primary health care. *Ann Fam Med* 2008;6:116-23.
- Rodríguez HP, Scoggins JF, von Glahn T, Zaslavsky AM, Safran DG. Attributing sources of variation in patients' experiences of ambulatory care. *Med Care* 2009;47:835-41.
- Sixma HJ, Spreeuwenberg PM, van der Pasch MA. Patient satisfaction with the general practitioner: a two-level analysis. *Med Care* 1998;36:212-29.
- Solomon LS, Zaslavsky AM, Landon BE, Cleary PD. Variation in patient-reported quality among health care organizations. *Health Care Financ Rev* 2002;23:85-100.
- Zaslavsky AM. Using hierarchical models to attribute sources of variation in consumer assessments of health care. *Stat Med* 2007;26:1885-900.
- Damman OC, Stubbe JH, Hendriks M, Arah OA, Spreeuwenberg P, Delnoij DM, et al. Using multilevel modeling to assess case-mix adjusters in consumer experience surveys in health care. *Med Care* 2009;47:496-503.
- Hargraves JL, Wilson IB, Zaslavsky A, James C, Walker JD, Rogers G, et al. Adjusting for patient characteristics when analyzing reports from patients about hospital care. *Med Care* 2001;39:635-41.
- Perneger TV. Adjustment for patient characteristics in satisfaction surveys. *Int J Qual Health Care* 2004;16:433-5.
- Salisbury C, Goodall S, Montgomery AA, Pickin DM, Edwards S, Sampson F, et al. Does Advanced Access improve access to primary health care? Questionnaire survey of patients. *Br J Gen Pract* 2007;57:615-21.
- National Primary Care Research & Development Centre. GPAQ: a patient questionnaire for general practice. 2009. www.gpaq.info.
- Salisbury C, Banks J, Goodall S, Baxter H, Montgomery A, Pope C, et al. An evaluation of advanced access in general practice. National Coordinating Centre for the Service Delivery and Organisation, 2007 (available at www.sdo.nihr.ac.uk/files/project/70-final-report.pdf).
- Leckie G, Goldstein H. The limitations of using school league tables to inform school choice. *J R Stat Soc Ser A* 2009;172:835-51.
- Hekkert KD, Cihangir S, Kleefstra SM, van den Berg B, Kool RB. Patient satisfaction revisited: a multilevel approach. *Soc Sci Med* 2009;69:68-75.
- Sheikh A. Ethnic minorities and their perceptions of the quality of primary care. *BMJ* 2009;339:b3797.

Accepted: 26 July 2010