

2011-10-27

# Factors associated with variability in the assessment of UK doctors' professionalism: analysis of survey results.

Campbell, JL

<http://hdl.handle.net/10026.1/4453>

---

10.1136/bmj.d6212

BMJ

---

*All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.*

# RESEARCH

## Factors associated with variability in the assessment of UK doctors' professionalism: analysis of survey results

 OPEN ACCESS

John L Campbell *professor*<sup>1</sup>, Martin Roberts *research fellow*<sup>1</sup>, Christine Wright *research fellow*<sup>1</sup>, Jacqueline Hill *research fellow*<sup>1</sup>, Michael Greco *associate professor*<sup>2</sup>, Matthew Taylor *project manager*<sup>2</sup>, Suzanne Richards *senior lecturer*<sup>1</sup>

<sup>1</sup>Peninsula College of Medicine and Dentistry, Exeter EX1 2LU, UK; <sup>2</sup>Client Focused Evaluation Programme (UK), Exeter

### Abstract

**Objectives** To investigate potential sources of systematic bias arising in the assessment of doctors' professionalism.

**Design** Linear regression modelling of cross sectional questionnaire survey data.

**Setting** 11 clinical practices in England and Wales.

**Participants** 1065 non-training grade doctors from various clinical specialties and settings, 17 031 of their colleagues, and 30 333 of their patients.

**Main outcome measures** Two measures of a doctor's professional performance using patient and colleague questionnaires from the United Kingdom's General Medical Council (GMC). We selected potential predictor variables from the characteristics of the doctors and of their patient and colleague assessors.

**Results** After we adjusted for characteristics of the doctor as well as characteristics of the patient sample, less favourable scores from patient feedback were independently predicted by doctors having obtained their primary medical degree from any non-European country; doctors practising as a psychiatrist; lower proportions of white patients providing feedback; lower proportions of patients rating their consultation as being very important; and lower proportions of patients reporting that they were seeing their usual doctor. Lower scores from colleague feedback were independently predicted by doctors having obtained their primary medical degree from countries outside the UK and South Asia; currently employed in a locum capacity; working as a general practitioner or psychiatrist; being employed in a staff grade, associate specialist, or other equivalent role; and with a lower proportion of colleagues reporting they had daily or weekly professional contact with the doctor. In fully adjusted models, the doctor's age, sex, and ethnic group were not independent predictors of patient or colleague feedback. Neither the age or sex profiles of the

patient or colleague samples were independent predictors of doctors' feedback scores, and nor was the ethnic group of colleague samples.

**Conclusions** Caution is necessary when considering patient and colleague feedback regarding doctors' professionalism. Multisource feedback undertaken for revalidation using the GMC patient and colleague questionnaires should, at least initially, be principally formative in nature.

### Background

In recent years, multisource feedback—the process of obtaining feedback from subordinates, peers, and supervisors—has been increasingly used in the business and health sectors to provide valuable information for workers about their performance, and as a means by which managers might stimulate improved performance. Previous research has suggested a complementary role for multisource feedback in performance appraisal.<sup>1</sup>

Regulatory bodies have the responsibility of monitoring the performance of doctors within their jurisdiction. In the United Kingdom, the General Medical Council (GMC) has proposed that doctors should undergo a process of revalidation, in which a clinician on the GMC register secures a continuing licence to practise on the grounds that they have demonstrated that they are “up to date and fit to practise” medicine.<sup>2</sup> All doctors on the GMC register were first issued with licences in 2009; revalidation is expected to be required from late 2012.<sup>3</sup>

Multisource feedback is seen as a source of valuable evidence to support or refute a doctor's application to revalidate. Multisource feedback is proposed as a central component in virtually all models of revalidation currently being considered by authoritative bodies in the UK<sup>2-5</sup> and elsewhere.<sup>6-7</sup> For doctors

Correspondence to: J L Campbell john.campbell@pms.ac.uk

Extra material supplied by the author (see <http://www.bmj.com/content/343/bmj.d6212/suppl/DC1>)

**Web appendix 1:** General Medical Council patient questionnaire

**Web appendix 2:** General Medical Council colleague questionnaire

who see patients, obtaining patient feedback is envisaged as part of the multisource feedback process.

In 2004, the GMC developed two survey instruments, proposed to support doctors in obtaining feedback from their patients and colleagues. Such approaches seek to assess whether a doctor actually does<sup>8</sup> deliver a high standard of professional practice by capturing information from workplace based assessors. We have previously provided evidence<sup>9</sup> regarding these instruments. After minor modifications to both instruments, we have also reported on the performance of those instruments in a large sample of doctors practising in various clinical settings in the UK.<sup>10</sup> The content of these instruments reflects the principles and values of medical professionalism as set out in the GMC's authoritative guidance for UK doctors.<sup>11</sup>

Statistical modelling of feedback about doctors' professionalism provides an opportunity to examine determinants of professional behaviour, inform processes of data collection, and explore potential predictors of the effect of assessors' and assessee's characteristics on performance scores. Many studies from the UK and elsewhere, including our own research, have used regression models to investigate the association between an individual doctor's performance and various characteristics of both the doctor being assessed and of those individuals providing assessment data.<sup>10 12-14</sup> These studies have modelled the scores provided by individual raters, giving insight into how assessor and assessee characteristics might affect the ratings. Such studies have highlighted, for example, that less than 15% of the variance in doctors' scores is accounted for by the extent of familiarity between observer and assessee.<sup>6 15</sup>

We aimed to investigate potential patient, colleague, and doctor related sources of systematic bias arising in the assessment of doctors' professionalism. In contrast with approaches based on the analysis of individual rater assessments, we have modelled the scores obtained by doctors as an average across a sample of their patient or colleague raters, allowing us to examine the effect on these average scores arising from variations in the profile of the rater groups and of the doctors themselves.

## Methods

Detailed methods have been reported elsewhere.<sup>10</sup> In summary, all non-training grade doctors from 11 sites in England and Wales were invited to take part between March 2008 and January 2011. The settings included four acute hospital trusts, an anaesthetics department from one acute trust, one mental health trust, four primary care organisations, and one independent sector organisation (that is, not part of the UK National Health Service). We aimed to recruit about 1000-1250 doctors across various practice settings and clinical specialties. We did not base this number on a formal sample size calculation, but rather aimed to obtain a sufficiently large sample to allow psychometric assessment of the data collection instruments.<sup>10</sup> We staged doctor recruitment and data collection at each site to avoid overburdening individual departments or practices. An internal communication was sent from the medical director or chief executive encouraging the doctors' participation. Doctors then received an information pack, containing a reply slip to indicate whether they wished to take part. We issued up to two reminders to non-responders.

Participating doctors were invited to identify up to 20 of their colleagues (half of whom were to be medically qualified) to take part in a secure online survey regarding the professionalism of the doctor. A paper alternative was available for colleague participants. Doctors were also invited to distribute, using administrative support if available, a paper based

post-consultation questionnaire and prepaid return envelope to 45 consecutive patients. The patient survey (web appendix 1) comprised nine core items relating to the doctor's performance, each scored using a five point scale. The colleague survey (web appendix 2) comprised 18 core items, which were also scored with a five point scale, using response options from "poor" (1) to "very good" (5) or from "strongly disagree" (1) to "strongly agree" (5) with higher scores indicating more positive ratings. All items included "don't know" or "not applicable" as relevant options. The personal characteristics of doctor participants were determined on the basis of self reports of a range of characteristics (table 1<sup>16</sup>).

After data return and cleaning, we calculated a summary patient score for each doctor, provided that at least 22 patient questionnaires had been returned, in line with our original instructions to participants to ensure adequate reliability.<sup>9</sup> We obtained the patient summary score by first calculating a mean score for each core item across patients where at least six patients had returned a valid score, and then calculating the mean of these item means where at least five of the possible nine core items means were available. We used a similar approach for feedback from colleagues, where at least eight colleagues had completed a questionnaire about the doctor's performance and more than half of the possible 18 core item means were available.

## Predicting doctor's patient and colleague scores

We used separate linear regression models to examine the association between a doctor's summary scores and a range of characteristics of the doctor, and of their patient and colleague samples—one model for patient scores, the other for colleague scores. Table 1 summarises the characteristics tested. We selected characteristics that had been identified as potentially important in pre-existing scientific literature.<sup>12-14 16-22</sup> We entered the identified predictor variables into the separate regression models for the patient and colleague scores. We used a significance threshold of  $P \leq 0.10$  to decide which characteristics of the doctors and rater groups should be included as potential independent predictors of the two mean summary scores in multiple regression models. If small subgroup sizes risked breaching anonymity (for example, in relation to the doctor's ethnic group), we combined categories of the relevant variables (table 1).

We regarded variables as significant independent predictors of the summary score if, after correcting for other variables in the model, the resulting P value was less than 0.05. We used bootstrapping to check the validity of the Wald based 95% confidence intervals, in view of the non-normality in the residuals, and we checked the regression models for sensitivity to the  $P \leq 0.10$  threshold for entering potential predictors. We calculated effect sizes for independent predictors in relation to the magnitude of the standard deviation of the respective patient or colleague score.<sup>23</sup>

## Results

Of 2454 invited doctors, 1065 (43%) agreed to take part, returning 30 333 patient questionnaires (mean 32.9 (standard deviation 10.8) per doctor) and 17 031 colleague questionnaires (16.1 (2.7)). Consent from patients and colleagues was indicated by the returning of the questionnaires. For 780 doctors who returned enough questionnaires to derive a patient score, the mean score was 4.80 (standard deviation 0.12, range 3.96-4.99); for 1050 doctors returning enough questionnaires to derive a

colleague score, the corresponding score were 4.63 (0.19, 3.57-4.96).

In univariate models, the doctor's sex, ethnic group, region of primary medical qualification, specialty group, and locum status were significantly associated with variation in patient scores (table 2<sup>1</sup>). The same variables, together with the doctor's age and current contractual role, were significantly associated with the colleague score. We therefore included these two sets of variables as potential predictors in the respective regression models.

In univariate models, seven patient related variables were significantly associated with the patient score (table 3<sup>2</sup>). Eight colleague related variables were significantly associated with the colleague score (table 4<sup>3</sup>). We also included these two sets of variables as potential predictors in the respective patient-doctor and colleague-doctor regression models.

## Predicting the patient summary mean score

Table 5<sup>4</sup> presents results for the final regression model for the patient summary score, based on data from 718 doctors who provided complete data on all relevant variables. The doctor's specialty group and region of primary medical qualification, together with the proportions of patients who were white, who regarded their visit as very important, or who were seeing their usual doctor were all independent predictors of patient scores. These predictors explained 21.0% of the variation in those scores. Doctors who had trained in South Asia or in jurisdictions outside the European Economic Area were likely to score lower on patient feedback than doctors trained in the UK. Psychiatrists were predicted to score lower than the general practitioner reference group. Increases in the proportions of patients who reported themselves as white, who regarded their visit as very important, or who reported seeing their usual doctor were all associated with increases in patient summary scores.

A large effect on patient feedback (effect  $>0.8^{23}$  × patient score standard deviation of 0.120) was evident for doctors from the psychiatry specialty group. After controlling for other variables in the analysis, psychiatrists were predicted to score 0.123 points lower than general practitioners and 0.143 points lower than doctors from other medical specialties. A large effect on patient score would also be expected to result from a 64% increase in the proportion of white patients in the sample, and from a 53% increase in the proportion of patients who regarded their visit as very important. Medium effects on patient scores were predicted for doctors who obtained their primary medical qualification in South Asia, and for a 63% increase in the proportion of patients reporting that they were seeing their usual doctor. Other effect sizes in respect of patient scores were small or not significant.

## Predicting the colleague related mean score

Table 6<sup>5</sup> shows results of the regression modelling for the colleague summary score, based on data from 949 doctors who provided complete data on all relevant variables. The doctor's specialty group, region of primary medical qualification, current contractual role, and locum status, together with the proportion of colleagues who reported daily or weekly contact with the doctor, were all independent predictors of the colleague summary score, together explaining 16.7% of the variation in those scores.

After controlling for other variables in the analysis, doctors trained outside the UK, except for those trained in South Asia, were likely to score lower than UK trained doctors. Consultants and general practitioners were likely to score 0.074 points higher

than doctors in other contractual roles, whereas doctors in locum posts were likely to score 0.093 points lower than those in permanent positions. Doctors in medical, surgical, and other specialty groups were predicted to score higher than the general practitioner reference group (by 0.091, 0.063, and 0.064 points, respectively). An increase in the proportion of colleagues reporting familiarity with the doctor's performance, based on daily or weekly contact with the doctor, was associated with an increase in colleague scores.

We did not see any large effects on colleague score arising from any of the variables examined. However, medium effects (effect  $>0.5^{23}$  × colleague score standard deviation of 0.194) were evident for doctors from medical, surgical, and other specialties compared with psychiatrists, and for a 70% increase in the proportion of colleagues reporting daily or weekly contact with the doctor during their period of familiarity.

## Discussion

### Summary of main findings

Using information obtained from the patients and colleagues of participating doctors, we found systematic variation in results of professionalism assessments among doctors working in a range of clinical settings and drawn from different clinical specialties. Some of the differences in doctors' scores after feedback from their patients and colleagues were attributable to differences between participating doctors in their personal and occupational characteristics. In addition, some of the differences in doctors' scores were attributable to variation between doctors in the characteristics and sociodemographic mix of their patients or colleagues in the feedback sample. These findings suggest that some doctors could be at risk of obtaining lower or higher scores based on sampling bias, rather than on the true variation between doctors in respect of their professional performance.

### Strengths and limitations

The research had several strengths. Firstly, our findings were based on a large sample of doctors with varying personal characteristics, drawn from several clinical settings and specialties. Furthermore, the patients and colleagues providing feedback varied widely in respect of their sociodemographic characteristics and in the nature of their relationship with the participating doctor. We have reported elsewhere<sup>10</sup> on the apparent acceptability of the multisource feedback process, as suggested by low levels of missing questionnaire data and high levels of assessor participation, and by the similar distribution in age and sex between doctors who were participants and those who were not. Finally, using regression models, we have identified a range of variables which independently predict doctors' scores after taking account of other variables in statistical models of doctors' professionalism. We have undertaken comprehensive modelling of the professionalism of fully trained doctors, taking account of both the characteristics of the doctor being assessed, and the characteristics of the sample of patient or colleague assessors.

In view of the current status of revalidation proposals in the UK, the study was, inevitably, based on a volunteer sample of doctors. We were reassured by the observed participation rate among all invited doctors (43%); although this rate was in excess of some other national level studies of doctors volunteering for multisource feedback,<sup>6 24</sup> we recognise that we might not have captured the full range of performance with respect to professionalism. In addition, to protect the anonymity of doctor participants, we incorporated data relating to some doctors from



small groups into larger groups before analysis. This was done, for example, for the small number of doctors reporting black ethnic status, whose feedback was incorporated with doctors from “other” ethnic groups.

## Doctor characteristics

Our models accounted for nine characteristics of the doctor whose professionalism was being assessed. Only two characteristics—the region of primary medical qualification and clinical specialty—were independent predictors of scores after patient feedback after also accounting for the mix of the patients providing feedback. In particular, doctors qualifying outside of Europe had lower patient feedback scores, as did psychiatrists.

Four doctor characteristics predicted colleague feedback: the region of primary medical qualification, clinical specialty, current contractual role, and locum status. Doctors who received lower feedback scores from their colleagues were those qualifying outside of the UK or South Asia, those working in locum posts, and those not working as a general practitioner or in a consultant role (such as doctors in associate specialist or staff grade roles). General practitioners and psychiatrists received reduced scores overall from their colleagues, compared with hospital based doctors.

It is perhaps gratifying that in modern day Britain with its tradition of equality legislation, the age, sex, and ethnic group of the doctor were not independent predictors of feedback scores from patients or colleagues. However, we found weak evidence suggesting that a doctor’s age and ethnic group were predictive of colleague feedback. Older doctors tended to have lower colleague feedback scores than younger doctors, and doctors of Asian ethnic origin had lower scores than those from white or other ethnic groups. To what extent these observations relate to true differences in performance as opposed to systematic variation in assessments based on non-clinical considerations is a matter of importance, and one which we cannot address in this study.

## Patient and colleague samples

We assessed the contribution of six characteristics of the patient feedback sample as potential predictors of overall patient summary scores in a model which also adjusted for the characteristics of the doctor being assessed. The proportions of white patient participants, patients identifying the reason for their consultation with the doctor as being very important, and patients reporting that they were seeing their usual doctor were independent predictors of more favourable patient scores. Neither the age or sex profiles of patient respondents, nor the proportion of respondents providing feedback as a proxy for the patient (for example, as a carer, or parent of a child), were predictors of patient feedback.

Using these data, we have been able to predict the effect of changes in the sociodemographic profile of the patient sample on doctors’ professionalism scores that might occur in doctors with a proportion of non-white patients that is higher than average. Our data identified that some doctors had no non-white patients in their sample, whereas for others, all of the patients providing feedback were from non-white ethnic groups. In addition, although many patients prefer continuity of care from their doctor,<sup>25 26</sup> fewer achieve this aspiration.<sup>27</sup> The dissonance between patients’ aspirations for continuity of care and their experience of care could, at least partly, be reflected in the reduced scores for professionalism attributed to doctors by patients who judged that they were not seeing their usual doctor.

Of eight characteristics of the colleague sample investigated as potential predictors of colleague scores, only one characteristic—the proportion of colleagues reporting that they had daily or weekly contact with the doctor being assessed during their period of familiarity with the doctor’s clinical practice—was a predictor of more favourable colleague feedback. Although this observation accords with findings reported by others,<sup>14 16-18</sup> Hall and colleagues<sup>6</sup> observed a negative effect of familiarity on ratings.

Seven other characteristics were not independent predictors of colleagues’ feedback scores, including the age, sex, and ethnic profile of the colleague sample; the proportion of colleague respondents who were in medical, other clinical, or administrative or managerial roles; the proportion of medically qualified colleagues who were in training grades; the proportion of colleagues who currently worked with the index doctor; and the proportion of colleagues returning their feedback using a paper questionnaire.

## Policy and practice implications

The UK regulator of medical practice, the GMC, has proposed major changes to the regulation of doctors, which are the most important changes to be introduced since the establishment of the GMC in 1858. Central to the proposed model for the revalidation of doctors are strengthened systems of appraisal, the appointment of “responsible officers” with a statutory role in “overseeing the evaluation of fitness to practise, and monitoring the conduct and performance of doctors,”<sup>28</sup> and the need for doctors to present evidence that they are “up to date and fit to practice.” Multisource feedback from colleagues, and, where appropriate, from patients, is seen as an important potential source of such evidence.

Although various clinical specialty groups could propose a range of evidence for an appraisal portfolio,<sup>29 30</sup> many doctors will probably seek, or be required to incorporate, feedback from patients and colleagues. Clinical specialty guidance should be based on authoritative evidence, recognising both the strengths and limitations of various approaches to providing evidence in relation to a doctor’s professionalism. The GMC has committed itself to issuing guidance to doctors on the use of questionnaires, and has noted the importance of using questionnaires that link to authoritative guidance on appropriate modern medical practice and that meet predetermined psychometric standards.<sup>2</sup>

Our data highlight the need for guidance for doctors in respect of identifying appropriate samples of colleagues and patients, and, importantly, the need for guidance for responsible officers in interpreting and responding to feedback on doctors’ professionalism. In particular, our data suggest that systematic bias might be responsible for at least some of the differences in the assessment of doctors’ performances, but this observation can only be confirmed by use of an objective measurement of professionalism. Adjusting scores to take account of the case mix might be an appropriate and potentially important response to these observations, facilitating interpretation of a doctor’s scores. Therefore, we advise careful consideration of the evidence which doctors might submit relating to their professionalism, and caution in developing judicious and appropriate responses to evidence which suggest a doctor’s performance to be unusual. Use of multisource feedback to support revalidation should at least initially be largely formative in nature and intent, and undertaken within the context of strengthened systems of appraisal.

**What is already known on this topic**

The GMC has proposed that UK doctors undergo revalidation to secure a continuing licence by demonstrating that they are “up to date and fit to practise” medicine

Multisource feedback from patients and colleagues is seen as an important source of evidence to support or refute a doctor's application to revalidate

**What this study adds**

Systematic bias may exist in the assessment of doctors' professionalism arising from the characteristics of the assessors giving feedback, and from the personal characteristics of the doctor being assessed

In the absence of a standardised measure of professionalism, doctors' assessment scores from multisource feedback should be interpreted carefully

Multisource feedback, for the purposes of supporting revalidation, should at least initially be largely formative in nature and intent, and undertaken within the context of strengthened systems of appraisal

We thank the doctors who contributed to this study, along with their patients, colleagues, and supporting administrative staff, for their cooperation and support; Tina Bealing and Louise Coleman (both of the Client Focused Evaluation Programme UK (CFEP-UK)) for their support of the project; Professor Martin Roland and Dr Obi Ukoumunne for their comments on earlier drafts of the research; and Helen Forster for proofreading the manuscript.

**Contributorship:** JC and MG conceived the study, and with SR and CW, developed the study design. JC is guarantor of the data. MG and MT oversaw data collection and initial processing. CW and JH monitored data collection. MR and JC undertook analysis. JC drafted the paper; all authors contributed to interpretation of the data and revision drafting of the text.

**Funding:** The study was funded by the UK GMC as an unrestricted research award. JC is an adviser to the GMC and has received only direct costs associated with presentation of this work. MG is a director of CFEP-UK and provided survey administration in respect of this research; MT was an employee of CFEP-UK at the time the research was undertaken. The study sponsor did not have any role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

**Competing interests:** All authors have completed the Unified Competing Interest form at [http://www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) (available on request from the corresponding author) and declare: the study was funded by the UK GMC as an unrestricted research award; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

**Ethical approval:** The study was considered by the Devon and Torbay NHS research ethics committee but judged not to require a formal ethics submission.

**Submission:** The submission of this paper conforms with the STROBE guidelines for cross sectional research studies (<http://www.strobe-statement.org/index.php?id=available-checklists>).

- Maylett T. 360-degree feedback revisited: the transition from development to appraisal. *Compens Benefits Rev* 2009;41:52-9.
- General Medical Council. Revalidation: the way ahead. GMC, 2010.
- General Medical Council. Supporting information for appraisal and revalidation. 2011. [www.gmc-uk.org/doctors/revalidation/revalidation\\_relicensing.asp](http://www.gmc-uk.org/doctors/revalidation/revalidation_relicensing.asp).
- Department of Health—Workforce Directorate. Medical revalidation—principles and next Steps. DH, 2008.
- Royal College of Physicians. Revalidation in practice. RCP, 2011.
- Hall W, Violato C, Lewkonja R, Lockyer J, Fidler H, Toews J, et al. Assessment of physician performance in Alberta: the physician achievement review. *CMAJ* 1999;161:52-7.
- Stern DT. Measuring medical professionalism. Oxford University Press, 2006.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;65:S63-7.

- Campbell JL, Richards SH, Dickens A, Greco M, Narayanan A, Brearley S. Assessing the professional performance of UK doctors: an evaluation of the utility of the General Medical Council patient and colleague questionnaires. *Qual Saf Health Care* 2008;17:187-93.
- Wright C, Richards S, Hill J, Roberts M, Norman G, Greco M, et al. Assessing professionalism: psychometric properties of the GMC Patient and Colleague Questionnaires. In submission, 2011.
- General Medical Council. Good medical practice. GMC, 2006.
- Franciosi M, Pellegrini F, De Berardis G, Belfiglio M, Di Nardo B, Greenfield S, et al. Correlates of satisfaction for the relationship with their physician in type 2 diabetic patients. *Diabetes Res Clin Pract* 2004;66:277-86.
- Salisbury C, Wallace M, Montgomery AA. Patients' experience and satisfaction in primary care: secondary analysis using multilevel modelling. *BMJ* 2010;341:c5004.
- Sargeant JM, Mann KV, Ferrier SN, Langille DB, Muirhead PD, Hayes VM, et al. Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: a pilot study. *Acad Med* 2003;78:S42-4.
- Lockyer J. Multisource feedback (360-degree evaluation). In: Holmboe ES, Hawkins RE, eds. *Practical guide to the evaluation of clinical competence*. Mosby, 2008:74-85.
- Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, Logerfo JP. Use of peer ratings to evaluate physician performance. *JAMA* 1993;269:1655-60.
- Lipner RS, Blank LL, Leas BF, Fortna GS. The value of patient and peer ratings in recertification. *Acad Med* 2002;77:S64-6.
- Wenrich MD, Carline JD, Giles LM, Ramsey PG. Ratings of the performances of practicing internists by hospital-based registered nurses. *Acad Med* 1993;68:680-7.
- Crossley J, McDonnell J, Cooper C, McAvoy P, Archer J, Davies H. Can a district hospital assess its doctors for re-licensure? *Med Educ* 2008;42:359-63.
- Mackillop LH, Crossley J, Vivekananda-Schmidt P, Wade W, Armitage M. A single generic multi-source feedback tool for revalidation of all UK career-grade doctors: does one size fit all? *Med Teach* 2011;33:e75-83.
- Weech-Maldonado R, Morales LS, Elliott M, Spritzer K, Marshall G, Hays RD. Race/ethnicity, language, and patients' assessments of care in Medicaid managed care. *Health Serv Res* 2003;38:789-808.
- Elliott MN, Zaslavsky AM, Goldstein E, Lehrman W, Hambarsoomian K, Beckett MK, et al. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res* 2009;44:501-18.
- Cohen J. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates, 1988.
- Violato C, Lockyer J, Fidler H. Multisource feedback: a method of assessing surgical practice. *BMJ* 2003;326:546-8.
- Gerard K, Salisbury C, Street D, Pope C, Baxter H. Is fast access to general practice all that should matter? A discrete choice experiment of patients' preferences. *J Health Serv Res Policy* 2008;13(suppl 2):3-10.
- Rademakers J, Delnoij D, de Boer D. Structure, process or outcome: which contributes most to patients' overall assessment of healthcare quality? *BMJ Qual Saf* 2011;20:326-31.
- Campbell SM, Kontopantelis E, Reeves D, Roland MO. Changes in patient experiences of primary care during health service reforms in England between 2003 and 2007. *Ann Fam Med* 2010;8:499-506.
- Department of Health. Responsible officers. DH, 2011.
- Bridgewater B, Cooper G, Livesey S, Kinsman R. Maintaining patients trust: modern medical professionalism 2011. Society for Cardiothoracic Surgery in Great Britain and Ireland. Dendrite Clinical Systems, 2011.
- Royal College of General Practitioners. RCGP guide to the revalidation of general practitioners, version 5. RCGP, 2010.

**Accepted:** 26 August 2011

**Cite this as:** *BMJ* 2011;343:d6212

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the license. See: <http://creativecommons.org/licenses/by-nc/2.0/> and <http://creativecommons.org/licenses/by-nc/2.0/legalcode>.

## Tables

**Table 1| Doctor, patient, and colleague sample variables**

Characteristic	Categories
<b>Index doctors (based on self report)</b>	
Sex	Male, female
Age	20-39, 40-49, 50-59, ≥60 years
Ethnic group*	White, Asian, other
Region of primary medical qualification†	UK, other European Economic Area jurisdictions, South Asia, other
Clinical specialty group‡	General practice, medical, surgical, psychiatry, mixed specialties (including group of other specialties)
Current contractual role	Consultant, non-consultant staff grade or associate specialist, general practitioner, other
Length of time in current contractual role	≤5, 6-15, 16-25, ≥26 years
Currently acting as a locum	Yes, no
Frequency of direct consultation with patients	Frequently, infrequently or occasionally, not at all
<b>Patients (to be presented as proportion (%) of patient sample)</b>	
Sex	Female
Age	<15 years, <21 years, >60 years
Ethnic group	White, Asian
Interaction with index doctor	Regarded visit to doctor as very important, seeing their usual doctor
Other	Questionnaire completed by a proxy
<b>Colleagues (to be presented as proportion (%) of colleague sample)</b>	
Sex	Female
Age	<30 years, ≥60 years
Ethnic group	White, Asian
Occupation	Doctors (including trainee doctors), trainee doctors, other clinical or health related role (for example, nursing), administration or managerial role
Interaction with index doctor	Currently works with index doctor, was or is in daily contact with index doctor, was or is in daily or weekly contact with index doctor
Survey mode	Returned paper version of questionnaire

\*Summarised from 16 groups originally.

†Summarised from list of 193 countries originally.

‡Summarised from list of 11 specialties originally.

**Table 2| Separate (unadjusted) regressions of patient summary score and colleague summary score on index doctor characteristics**

Index doctor characteristic	P	
	Patient score	Colleague score
Age	0.509	<0.001*
Sex	0.022*	0.055*
Ethnic group	<0.001*	<0.001*
Region of primary medical qualification	<0.001*	<0.001*
Clinical specialty group	<0.001*	<0.001*
Current contractual role	0.191	<0.001*
Time in current contractual role	0.207	0.220
Locum status	0.002*	<0.001*
Current role involves direct consultation with patients	0.641	0.774

\*P value of  $\leq 0.10$  used to identify potential predictor variables for multiple regression modelling.



**Table 3| Descriptive statistics for patient related variables and separate unadjusted regressions on patient summary score**

Proportion of characteristic in patient sample	Mean (SD)*	Range*	P
Female	61 (19)	0-100	0.035†
<15 years old	7 (15)	0-89	0.798
<21 years old	10 (17)	0-96	0.077†
>60 years old	39 (25)	0-100	0.001†
White ethnic group	95 (10)	0-100	<0.001†
Asian ethnic group‡	3 (7)	0-100	<0.001†
Regarded visit to doctor as very important	67 (14)	11-100	<0.001†
Seeing their usual doctor	57 (27)	0-100	<0.001†
Questionnaire completed by a proxy	15 (20)	0-100	0.679
Returned the questionnaire by post	50 (41)	0-100	0.176

\*Data are percentage values.

†P value of ≤0.10 used to identify potential predictor variables for multiple regression modelling.

‡Variable subsequently dropped from multiple regression model due to collinearity.

**Table 4| Descriptive statistics for colleague related variables and separate unadjusted regressions on colleague summary score**

Proportion of characteristic in colleague sample	Mean (SD)*	Range*	P
Female	58 (18)	0 to 100	0.875
<30 years old	4 (6)	0 to 40	0.096†
≥60 years old	8 (10)	0 to 60	0.004†
White ethnic group	87 (14)	26 to 100	<0.001†
Asian ethnic group‡	9 (11)	0 to 70	<0.001†
Doctors (including trainee doctors)	50 (14)	0 to 100	0.563
Trainee doctors§	4 (6)	0 to 42	0.818
Other clinical or health related roles	29 (12)	0 to 85	0.283
Administration or managerial roles	20 (11)	0 to 78	0.159
Currently works with doctor	81 (13)	13 to 100	<0.001†
Was or is in daily contact with doctor‡	46 (23)	0 to 100	0.063†
Was or is in daily or weekly contact with doctor	83 (17)	6 to 100	<0.001†
Returned paper version of questionnaire	15 (25)	0 to 100	0.074†
Doctor colleagues who are trainees¶	8 (13)	0 to 100	0.846

\*Data are percentage values.

†P value of ≤0.10 used to identify potential predictor variables for multiple regression modelling.

‡Variables subsequently dropped from multiple regression model due to colinearity.

§Refers to percentage of trainee doctors in the entire sample.

¶Refers to percentage of doctors in the sample who are trainees.

Table 5| Fully adjusted linear regression results for patient score

	No of doctors in subgroup*	Change in patient score (95% CI)†‡	P
<b>Doctor characteristics</b>			
Sex			
Male	470	Reference	0.084
Female	248	0.016 (–0.002 to 0.035)	
Ethnic group			
White	571	Reference	0.975
Asian	109	0.001 (–0.039 to 0.040)	
Other	38	–0.004 (–0.043 to 0.035)	
Region of primary medical qualification			
UK	538	Reference	<0.001
European Economic Area (non-UK)	37	–0.031 (–0.068 to 0.006)	
South Asia	77	–0.077 (–0.123 to –0.030)	
Other	66	–0.056 (–0.088 to –0.025)	
Medical specialty group			
General practice	333	Reference	<0.001
Medical	201	0.020 (–0.005 to 0.044)	
Surgical	124	0.011 (–0.015 to 0.037)	
Psychiatry	24	–0.123 (–0.170 to –0.077)	
Other	36	0.015 (–0.024 to 0.054)	
Locum status			
Non-locum	696	Reference	0.358
Locum	22	–0.022 (–0.070 to 0.025)	
<b>Patient sample characteristics§</b>			
Female	–	0.004 (–0.001 to 0.009)	0.089
<21 years old	–	–0.001 (–0.007 to 0.005)	0.816
>60 years old	–	0.002 (–0.002 to 0.007)	0.336
White ethnic group	–	0.015 (0.006 to 0.024)	0.001
Regarded visit to doctor as very important	–	0.018 (0.011 to 0.026)	<0.001
Seeing their usual doctor	–	0.010 (0.006 to 0.013)	<0.001

\*Total sample size=718 doctors.

†Regression coefficients for doctor characteristics represent change in patient score (on 1-5 scale) expected for doctors in particular subgroups relative to the reference subgroup. Thus, psychiatrists would be expected to score 0.123 points lower than general practitioners. Most changes are less than one patient score standard deviation of 0.120.

‡Wald based confidence intervals.

§Regression coefficients for patient sample characteristics are the change in patient score expected for a 10% increase in the variable.

Table 6| Fully adjusted linear regression results for colleague score

	No of doctors in subgroup*	Change in colleague score (95% CI)†‡	P
<b>Index doctor characteristics</b>			
<b>Age</b>			
20-39 years	182	Reference	0.054
40-49 years	429	0.013 (–0.019 to 0.045)	
50-59 years	260	–0.015 (–0.052 to 0.021)	
≥60 years	78	–0.048 (–0.102 to 0.006)	
<b>Sex</b>			
Male	619	Reference	0.559
Female	330	0.008 (–0.018 to 0.033)	
<b>Ethnic group</b>			
White	750	Reference	0.051
Asian	146	–0.075 (–0.135 to –0.014)	
Other	53	–0.020 (–0.076 to 0.037)	
<b>Region of primary medical qualification</b>			
UK	707	Reference	0.002
European Economic Area (non-UK)	49	–0.089 (–0.142 to –0.036)	
South Asia	107	–0.037 (–0.107 to 0.033)	
Other	86	–0.058 (–0.105 to –0.012)	
<b>Medical specialty group</b>			
General practice	355	Reference	<0.001
Medical	320	0.091 (0.060 to 0.122)	
Surgical	169	0.063 (0.027 to 0.100)	
Psychiatry	53	–0.039 (–0.094 to 0.016)	
Other	52	0.064 (0.010 to 0.118)	
<b>Current contractual role</b>			
Consultant or general practitioner	825	Reference	<0.001
Staff grade, associate specialist, or other equivalent role	124	–0.074 (–0.110 to –0.037)	
<b>Locum status</b>			
Non-locum	926	Reference	0.017
Locum	23	–0.093 (–0.170 to –0.017)	
<b>Colleague sample characteristics§</b>			
<30 years old	–	0.016 (–0.003 to 0.035)	0.102
≥60 years old	–	0.006 (–0.009 to 0.020)	0.451
White ethnic group	–	0.001 (–0.009 to 0.012)	0.834
Currently works with index doctor	–	0.000 (–0.011 to 0.010)	0.946
Was or is in daily or weekly contact with index doctor	–	0.014 (0.006 to 0.022)	0.001
Returned paper version of questionnaire	–	0.003 (–0.002 to 0.009)	0.243

\*Total sample size=949 doctors.

†Regression coefficients for doctor characteristics represent change in colleague score (on 1-5 scale) expected for doctors in particular subgroups relative to the reference subgroup. Thus, locum doctors would be expected to score 0.093 points lower than non-locum doctors. All changes are less than one colleague score standard deviation of 0.194.

‡Wald based confidence intervals.

§Regression coefficients for colleague sample characteristics are the change in colleague score expected for a 10% increase in the variable.