

The Evaluation of Physicians' Communication Skills From Multiple Perspectives

Jenni Burt, PhD¹

Gary Abel, PhD²

Marc N. Elliott, PhD³

Natasba Elmore, MSc¹

Jennifer Newbould, PhD⁴

Antoinette Davey, MPhil²

Nadia Llanwarne, MPhil⁵

Inocencio Maramba, MSc²

Charlotte Paddison, PhD⁵

John Campbell, MD, FRCGP²

Martin Roland, DM, FRCGP⁵

¹The Healthcare Improvement Studies Institute (THIS Institute), University of Cambridge, Cambridge Biomedical Campus, Cambridge, United Kingdom

²University of Exeter Medical School, St Luke's Campus, Exeter, United Kingdom

³RAND Corporation, Santa Monica, California

⁴RAND Europe, Cambridge, United Kingdom

⁵Cambridge Centre for Health Services Research, University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom



Conflicts of interest: M.R. and J.C. have acted as advisors to Ipsos MORI, the Department of Health and subsequently NHS England on the development of the English GP Patient Survey. J.B. currently acts as an advisor to NHS England on the GP Patient Survey. No other authors report a conflict of interest.

CORRESPONDING AUTHOR

Jenni Burt, PhD

The Healthcare Improvement Studies Institute (THIS Institute)
University of Cambridge
Cambridge Biomedical Campus
Clifford Allbutt Building
Cambridge CB2 0AH UK
jenni.burt@thisinstitute.cam.ac.uk

ABSTRACT

PURPOSE To examine how family physicians', patients', and trained clinical raters' assessments of physician-patient communication compare by analysis of individual appointments.

METHODS Analysis of survey data from patients attending face-to-face appointments with 45 family physicians at 13 practices in England. Immediately post-appointment, patients and physicians independently completed a questionnaire including 7 items assessing communication quality. A sample of videotaped appointments was assessed by trained clinical raters, using the same 7 communication items. Patient, physician, and rater communication scores were compared using correlation coefficients.

RESULTS Included were 503 physician-patient pairs; of those, 55 appointments were also evaluated by trained clinical raters. Physicians scored themselves, on average, lower than patients (mean physician score 74.5; mean patient score 94.4); 63.4% (319) of patient-reported scores were the maximum of 100. The mean of rater scores from 55 appointments was 57.3. There was a near-zero correlation coefficient between physician-reported and patient-reported communication scores (0.009, $P = .854$), and between physician-reported and trained rater-reported communication scores (-0.006, $P = .69$). There was a moderate and statistically significant association, however, between patient and trained-rater scores (0.35, $P = .042$).

CONCLUSIONS The lack of correlation between physician scores and those of others indicates that physicians' perceptions of good communication during their appointments may differ from those of external peer raters and patients. Physicians may not be aware of how patients experience their communication practices; peer assessment of communication skills is an important approach in identifying areas for improvement.

Ann Fam Med 2018;16:330-337. <https://doi.org/10.1370/afm.2241>.

INTRODUCTION

Patient-centered communication is fundamental to the practice of family medicine.^{1,2} While good communication itself is an important outcome, it is associated with benefits such as improvement of clinical outcomes, reduction in medical errors, and facilitation of self-management and preventive behaviors.³⁻¹¹ Internationally, the evaluation of physicians' communication skills is increasing as part of efforts to improve the quality of health care.¹²⁻¹⁴ Approaches to evaluating and benchmarking standards of communication have typically relied on patient experience surveys, the results of which are often made public.^{15,16} At the level of the individual, physicians may need to reflect on their own performance alongside ratings from peers, coworkers, and patients as part of both regulation and continuing professional development.¹⁷⁻²⁰ For example, in the UK, the General Medical Council requires all doctors to complete 360-degree evaluation of the care they provide, with patient and colleague feedback used as supporting information for the renewal of their license to practice.²¹

Confidence in the instruments used to assess—and commonly compare—performance is essential if they are to contribute meaningfully to quality assurance.²² Extensive work on the reliability and validity of patient questionnaires has been conducted.²³⁻²⁸ Despite this, research shows that doctors often struggle to trust, make sense of, and subsequently respond to, feedback from patient surveys.²⁹⁻³¹ In fact, evidence from evaluations of performance (aggregated across a series of appointments) suggests that physicians tend to rate themselves more negatively than patients or peers.³²⁻³³ Indeed, physicians' perceptions of their own competence are frequently out of line with external assessments, as patients tend to give particularly favorable assessments of care in comparison to the physician self-assessments.³⁴⁻³⁷ The greatest divergence between self-assessments of physicians and others, however, is with those physicians who are, by external evaluation, the least skilled but the most confident in their abilities (a phenomenon not confined to physicians alone).^{34,38,39}

To date, research in the area of reliability and validity of patient questionnaires has focused on the evaluation of overall performance assessed across a series of appointments.^{18,19} We compared physician, patient, and rater assessments of communication for individual appointments to discover where discrepancies in assessment of care originate and to learn about physicians' insight into patients' perceptions of care during a single encounter. While we considered differences in the distribution of scores given by raters, patients, and physicians, our main focus was on correlations of scores at the appointment level. The correlations were considered more important for assessing the extent to which physicians are able to distinguish (1) appointments that more fully met communications standards from those appointments that did so to a lesser extent, and (2) appointments that resulted in better patient experiences from those that resulted in worse patient experiences. The correlation of patient and rater scores is also of interest as it illuminates the extent to which use of communication best practices may improve patient experience.

METHODS

We present an analysis of data collected in a study conducted in family practices in England in 2 broad geographic areas (Devon, Cornwall, Bristol, and Somerset; and Cambridgeshire, Bedford, Luton, and North London). Approval for the study was obtained from the National Research Ethics Service Committee East of England – Hertfordshire on 11 October 2011 (ref: 11/EE/0353).

Sampling and Practice Recruitment

Practices were eligible if they (1) had more than 1 family physician (physician) working a minimum of 2 days a week in direct clinical contact with patients, and (2) had low scores on physician-patient communication items used in the 2009-2010 national GP (general practitioner) Patient Survey. Low scores were defined as below the lower quartile for mean communication score, adjusted for patient case-mix (age, sex, ethnicity, self-rated health, and an indicator of area-level deprivation/disadvantage).⁴⁰ This study was part of a research program concerned with understanding the full range of patient experiences of communication, from poor to good.⁴¹ In England, however, 94% of patients score all questions addressing GP communication during appointments as good or very good in the GP Patient Survey: therefore, we specifically sought low-scoring practices to maximize the chance of some appointments within the practice being given low patient ratings for communication. We approached eligible practices within the study areas until we had recruited 13 practices: some practices were known to us from participation in a previous study in the program.⁴⁰

Patient Recruitment

Data collection took place between August 2012 and July 2014, with recruitment of 1 or 2 physicians at a time in each practice. As the primary component of the study involved video-recording the encounter (reported elsewhere⁴²), we based researchers in the practice to recruit patients into the study. The research team approached adult patients on their arrival in the practice for a face-to-face appointment with a participating physician. Patients received a summary, a detailed information sheet, and a consent form. A member of the research team discussed these with each participating patient in order to obtain informed consent.

Patient and Physician Ratings

Immediately following the appointment, the patient was asked to complete a short questionnaire. The questionnaire included a set of 7 items taken from the national GP Patient Survey to assess physician-patient communication (Table 1) and basic sociodemographic questions. Also, following the appointment, physicians answered the same 7 items about their own communication performance in that encounter. From these, we calculated separate scores of communication during the appointment, from the patient responses and from the physician responses. In line with previous work, each was calculated by linearly rescaling responses from 0 to 100 and calculating the mean of all informative responses where 4 or more informative answers were

given.^{40,43} Responses of “doesn’t apply” were considered uninformative and excluded.

Trained Clinical Rater Ratings

In addition to physician self ratings and patient ratings, 56 of the consultations were selected for rating by experienced, trained clinical raters (all family physicians). The selection of appointments was made on the basis of patient ratings of communication, with the aim of maximizing the variation of patient-reported communication quality. To increase reliability, 4 raters scored each appointment, using both the Global Consultation Rating Scale⁴⁴ and the same set of 7 items taken from the GP Patient Survey used by patients and physicians. Full details of the rating process were reported in a previous publication, which showed a weak correlation between patient ratings of physician communication and trained raters scores using the Global Consultation Rating Scale.⁴² In this analysis, we made use of the items derived from the GP Patient Survey, calculating scores as described above. Each of the raters scored appointments in a different random order to minimize any order effects (using simple randomization) and the same raters were used for all appointments. The mean of the scores from the 4 raters was calculated for each appointment.

Statistical Analyses

We calculated correlation coefficients comparing physician and patient scores for the full sample and physician, patient, and rater scores for a subsample. To evaluate the within-physician association between patient and physician scores, we used a mixed linear regression with a random effect (intercept) for each physician on the full sample. This model accounts for the fact that some physicians may be more generous or more critical than others, and thus assessed

whether individual physicians’ scores for particular appointments increased when patients also rated them higher. This mixed model was performed initially with a single fixed effect (patient-reported scores) and subsequently adjusted for patient demographics (age, sex, ethnicity, and self-rated health) to account for the fact that some types of patients were more likely to give positive ratings. Another model performed included physician sex, whether they were UK qualified, and the years since they qualified, to adjust for any differences not captured by the random effect for physician. Standardized regression coefficients (betas) are reported. These are directly comparable to (and in the case of models with a single exposure, equal to) correlation coefficients. Because of potential concerns over normality assumptions, bootstrapping was used in all analyses with 500 bootstrap samples. To account for the nonindependence of observations due to physicians being represented more than once, we performed the bootstrap sampling clustered by physician. All analysis was carried out using Stata V13.1 (StataCorp LP).

RESULTS

A total of 908 patients had face-to-face appointments with 45 participating physicians during periods of patient recruitment. Of these, 167 (18.4%) were ineligible (mostly children) and, of the remainder, 529 completed a questionnaire (71.4% response rate). An additional 26 (4.9%) appointments were excluded due to missing data, leaving 503 physician-patient appointment pairings in the data set (Supplemental Figure 1, available at <http://www.annfam.org/content/16/4/330/suppl/DC1/>). Table 2 shows self-reported demographic characteristics of patients. For 4 physicians, data on sex, country of qualification,

Table 1. Physician-Patient Communication Items

Instructions given to patients: “Thinking about the consultation which took place today, how good was the doctor at each of the following?^a Please put an X in 1 box for each row.”

	Very good	Good	Neither good nor poor	Poor	Very poor	Doesn’t apply ^b
Giving you enough time	<input type="checkbox"/>					
Asking about your symptoms	<input type="checkbox"/>					
Listening to you	<input type="checkbox"/>					
Explaining tests and treatments	<input type="checkbox"/>					
Involving you in decisions about your care	<input type="checkbox"/>					
Treating you with care and concern	<input type="checkbox"/>					
Taking your problems seriously	<input type="checkbox"/>					

^a Amended for physicians to read “How good were you at each of the following?” Items were also reworded, for example, “Giving the patient enough time.” Raters were asked, “Thinking about this consultation, how good was the ...?”

^b Considered to be uninformative for the purposes of our analysis.

Table 2. Self-Reported Demographics for Patients Who Completed a Questionnaire

Characteristic	N (%)	Characteristic	N (%)
Sex		Self-rated health	
Male	199 (39.56)	Excellent	49 (9.74)
Female	304 (60.44)	Very good	163 (32.41)
Age, y		Good	179 (35.59)
18-24	36 (7.16)	Fair	78 (15.51)
25-34	76 (15.11)	Poor	34 (6.76)
35-44	61 (12.13)	Ethnicity	
45-54	78 (15.51)	White	458 (91.05)
55-64	83 (16.50)	Mixed	4 (0.80)
65-74	93 (18.89)	Asian or Asian British	15 (2.98)
75-84	59 (11.73)	Black or Black British	21 (4.17)
85+	15 (2.98)	Chinese	4 (0.80)
		Other	1 (0.20)

and date of qualification were not available. Of the 56 appointments selected for evaluation by raters, 55 (98%) had complete physician and patient scores and the subsample analysis was restricted to these appointments. The individual rater scores for the 55 consultations were strongly correlated with each other (pair-wise Spearman correlation coefficients varied between 0.54 and 0.67, $P < .0001$ for all, see Supplemental Table 1, available at <http://www.annfam.org/content/16/4/330/suppl/DC1/>) giving confidence that the scale was being used consistently and that using the mean of the 4 rater scores was appropriate.

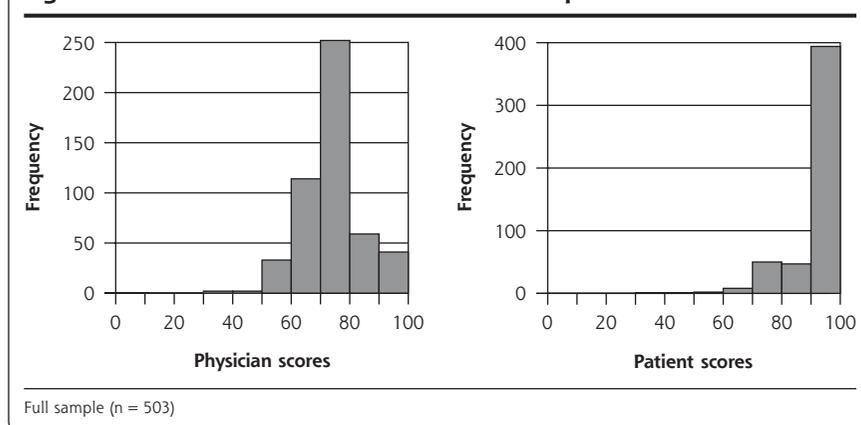
Physician and Patient Comparison

Figure 1 shows the distribution of physician-reported and patient-reported scores for the full sample (503 appointments). Physician scores of their performance were fairly symmetrically distributed and ranged from 39.3 to 100 (mean 74.5), with only 5.4% (27) of appointments being given the maximum score of 100. In contrast, the distribution of patient-reported scores is highly skewed, with 63.4% (319) of patients giving the maximum score of 100 (range 32.1 to 100, mean 94.4). A scatter plot comparing physician-reported scores with patient-reported scores of the same appointment is shown in Figure 2. The skewed nature of patient scores is evident in this figure, which also shows that, while physicians do not often score themselves lower than

50, on average they gave themselves much lower scores than patients. The lack of any clear relationship in Figure 2 is reflected in the very low correlation coefficient shown in Table 3, again with no evidence of an association ($P = .854$). The lack of association persists when considering within-physician associations and when further adjusting for patient demographics (Table 3). Additional adjustment for physician factors had no meaningful impact on the regression coefficient or P value for physician self-rating.

Physician, Patient, and Rater Comparison

Figure 3 shows the distribution of physician-reported, patient-reported, and rater-reported scores for the 55 appointment subsample. The bi-modal distribution of patient scores reflects the way appointments were sampled, while the physician self-rated scores were distributed similarly to the full sample. The raters scored appointments over a wider range than either patients or physicians, from 23.2 to 87.5 (mean 57.3), and their scores were less skewed than those of patients. Figure 4 shows scatter plots comparing the 3 sets of ratings. Similar to the full data set shown in Figure 2, there is no association between physician scores and patient scores in the subset of appointments evaluated by raters. Furthermore, there is no association between physician scores and the scores of raters, although there is a tendency for patient scores to be higher when the rater scores were also higher. These relationships are reflected in the correlation coefficients of 0.015 ($P = .91$) for physicians and patients, -0.006 ($P = .69$) for physicians and raters, and 0.35 ($P = .042$) for patients and raters. The only pair with any statistically significant and nontrivial association is between the scores of patients and raters.

Figure 1. Distribution of scores for the full sample.

DISCUSSION

In this examination of family physicians', patients', and trained clinical raters' assessments of physician-patient communication during individual appointments, we

found no correlation between physician and patient scores or between physician and rater scores, and a moderate correlation between patient and rater scores.

Our results suggest that family physicians draw on different constructs of good communication compared with patients and trained clinical raters, when asked to complete the same evaluation items. Previous research has documented a mismatch between physicians' assessments of patient expectations, their subsequent communication behaviors, and patient perceptions of these behaviors, most notably in pediatric appointments.³⁴⁻³⁷ Our findings suggest that a divergence between physician and patient expectations of communication practices may be common in primary care. Therefore, physicians' self-perceptions alone may be of limited value for identifying aspects of their patient-centered communication practices which could be strengthened or improved. Raters are more likely to share patient perceptions of what good communication looks like. Additionally, raters may pinpoint aspects of physicians' communication behaviors which are not perceived by patients, or at least not reported in a post-consultation survey.³³ Multisource feedback for the assessment of physician performance is now an established tool for

Figure 2. Scatterplot illustrating the association between physician and patient scores.

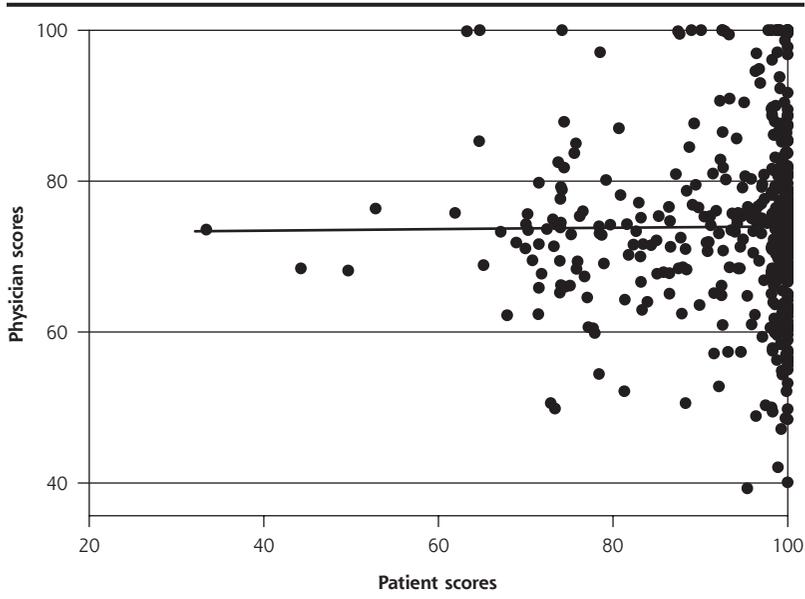
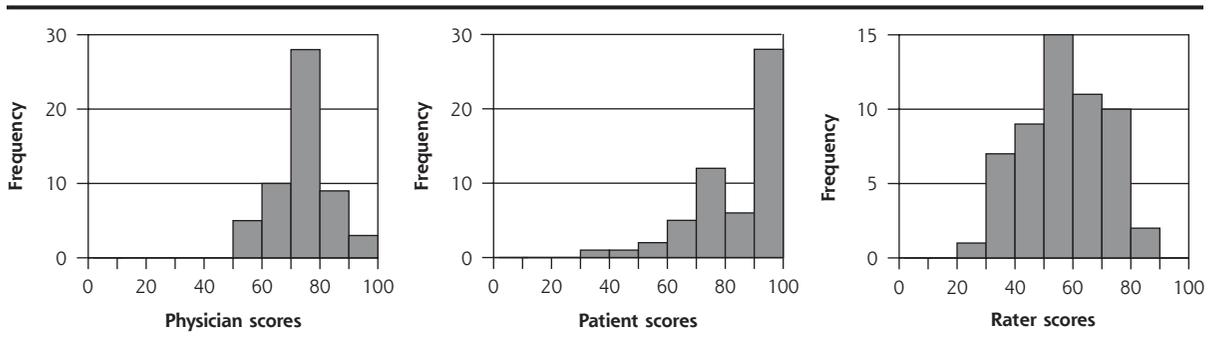


Table 3. Standardized Regression Coefficients for Physician and Patient Scores (n = 503)

Association	Standardised Regression Coefficients (95% CI)	P value
Overall association	0.009 (-0.086 to 0.104)	.854
Within-physician association	0.025 (-0.060 to 0.110)	.565
Within-physician association adjusted for patient sociodemographics	0.023 (-0.064 to 0.110)	.608
Within-physician association adjusted for patient sociodemographics and physician factors ^a	0.051 (-0.044 to 0.146)	.291

^aRestricted to 451 consultations where physician data were available.

Figure 3. Distribution of scores for the subsample.



Subsample rated by trained raters (n = 55)

evaluating the quality of care, with increasing evidence of impact on physician behaviors.^{45,46} Our study provides further evidence for the importance of external assessment of physicians' communication skills by trained peers as a first step in improving the standard of physician-patient communication.

The differences we observed in the distribution of scores used by raters, patients, and physicians are of

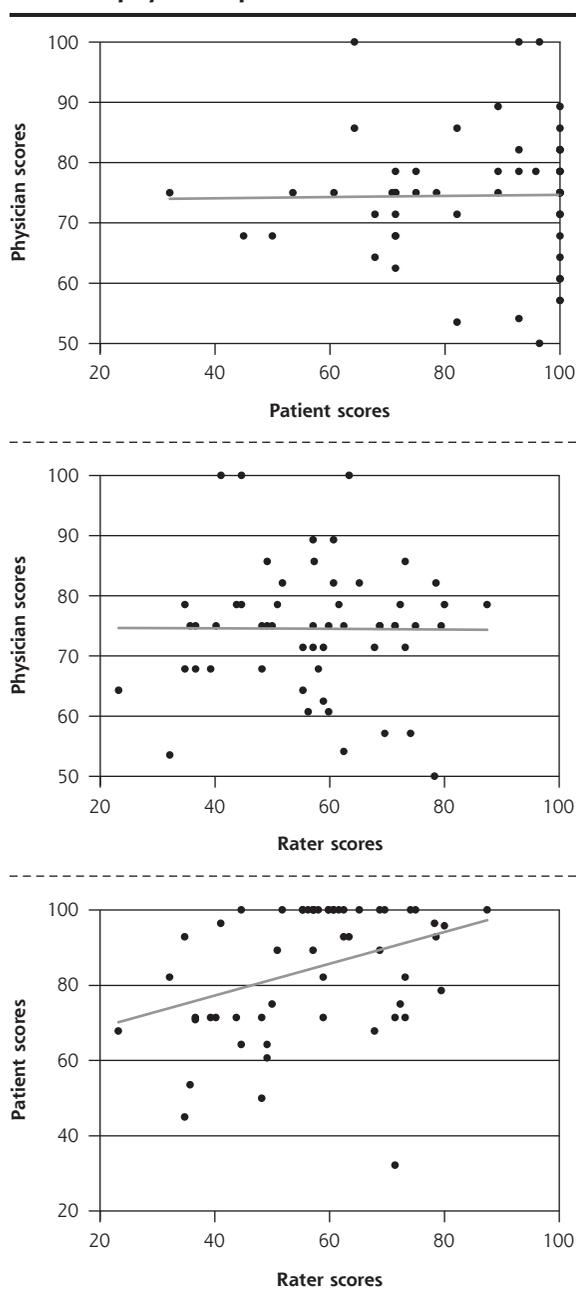
interest, although they must be interpreted with some caution. Patients provided more generous scores, on average, than raters or physicians. High patient scores reflect, in part, the fact that some patients are inhibited about identifying poor communication on patient experience questionnaires.⁴⁷ The reluctance of some patients to report poor experiences is likely to result in weaker correlations between patient and rater assessments of communication than would otherwise occur. In aggregate, patient ratings are able to distinguish the quality of physician performance overall.^{48,49} Given the different range of scores used by each group (patients, physicians, raters) on the same response scale, however, we suggest that patient experience scores are best interpreted as a relative measure of the patient experience, rather than being interpreted on an absolute scale.⁴² This further supports the need for external peer assessment of communication skills, as patient feedback alone is unlikely to identify specific needs for support and training in this area.

Our study has a number of limitations. We selected practices to increase the likelihood of identifying appointments with lower patient scores. Within each practice, not all physicians took part. If physicians who participated were more skilled at communicating with patients, we may have reduced the variation in quality of communication in our sample, thus reducing study power and the strength of the observed correlations.

We asked physicians to assess their communication performance immediately after each appointment, when time may have been short. Thus our findings may not be generalizable to other forms of self-reflection where more time is taken, for example, in review of video-recorded appointments. On the other hand, our method of data collection may be representative of the informal self-evaluation that routinely occurs among physicians. We additionally note that we did not assess the compliance of each participating physician to our request to complete an assessment after every appointment. While we collected assessments at the end of each surgery, reliability may have been reduced if physicians completed assessments in batches following a series of appointments. Patients completed questionnaires immediately following their appointment, usually in the practice waiting area. Social desirability bias may have increased the likelihood of patients giving positive assessments of care. Additionally, most patients in the study self-reported as white and our findings may not generalize well to patients of different racial and ethnic backgrounds.

Patient feedback is, and should remain, a central component of assessments of the quality of care. Our findings, however, support the role of trained peer assessors in examining the communication practices of

Figure 4. Scatterplots illustrating associations between physician, patient, and rater scores.



Note: The gray lines are lines of best fit.

physicians in any multisource assessment investigating standards of care. We would further suggest that the presentation of feedback from such assessments should include support for physicians to better attune themselves to the perceptions and communication needs of their patients.

To read or post commentaries in response to this article, see it online at <http://www.AnnFamMed.org/content/16/4/330>.

Key words: physician-patient relations; health care surveys; quality of health care; patient satisfaction; patient experience; physician-patient communication; health care quality measurement

Submitted September 1, 2017; submitted, revised, January 30, 2018; accepted February 27, 2018.

Funding support: This work was funded by a National Institute for Health Research Programme Grant for Applied Research (NIHR PGfAR) program (RP-PG-0608-10050).

Department of Health disclaimer: The views expressed are those of the author(s) and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health.

Acknowledgments: We would like to thank the patients, practice managers, family physicians, and other staff of the general practices who kindly agreed to participate in this study and without whom the study would not have been possible. Particular acknowledgment goes to our 4 trained clinical raters for their contribution to this work, and to James Brimicombe, our data manager, who developed the online rating system. We would also like to thank the Improve Advisory Group for their input and support throughout this study.

Supplementary Materials: Available at <http://www.AnnFamMed.org/content/16/4/330/suppl/DC1/>.

References

- Buetow SA. What do general practitioners and their patients want from general practice and are they receiving it? A framework. *Soc Sci Med*. 1995;40(2):213-221.
- Wensing M, Jung HP, Mainz J, Olesen F, Grol R. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. *Soc Sci Med*. 1998;47(10):1573-1588.
- Anhang Price R, Elliott MN, Zaslavsky AM. Valuing patient experience as a unique and intrinsically important aspect of health care quality. *JAMA Surg*. 2013;148(10):985-986.
- Anhang Price R, Elliott MN, Zaslavsky AM, et al. Examining the role of patient experience surveys in measuring health care quality. *Med Care Res Rev*. 2014;71(5):522-554.
- Stewart MA. Effective physician-patient communication and health outcomes: a review. *CMAJ*. 1995;152(9):1423-1433.
- Kuzel AJ, Woolf SH, Gilchrist VJ, et al. Patient reports of preventable problems and harms in primary health care. *Ann Fam Med*. 2004;2(4):333-340.
- Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open*. 2013;3(1):e001570.
- Kennedy M, Denise M, Fasolino M, et al. Improving the patient experience through provider communication skills building. *Patient Exp J*. 2014;1(1):56-60.
- Kennedy A, Gask L, Rogers A. Training professionals to engage with and promote self-management. *Health Educ Res*. 2005;20(5):567-578.
- Cohen D, Longo MF, Hood K, Edwards A, Elwyn G. Resource effects of training general practitioners in risk communication skills and shared decision making competences. *J Eval Clin Pract*. 2004;10(3):439-445.
- Kroenke K. A practical and evidence-based approach to common symptoms: a narrative review. *Ann Intern Med*. 2014;161(8):579-586.
- Health Research Institute. *Scoring Healthcare: Navigating Customer Experience Ratings*. New York, NY: PricewaterhouseCoopers; 2013.
- Department of Health. *Hard Truths: The Journey to Putting Patients First*. London, UK: The Stationery Office; 2014.
- Ahmed F, Burt J, Roland M. Measuring patient experience: concepts and methods. *Patient*. 2014;7(3):235-241.
- Agency for Healthcare Research and Quality (AHRQ). CAHPS Clinician and Group Surveys. <https://cahps.ahrq.gov/surveys-guidance/cg/index.html>. Published 2018. Accessed Jan 27, 2018.
- Ipsos MORI. GP Patient Survey. <https://gp-patient.co.uk/>. Published 2018. Accessed Jan 27, 2018.
- Lockyer J. Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof*. 2003;23(1):4-12.
- Overeem K, Wollersheim HC, Arah OA, Cruisberg JK, Grol RP, Lombarts KM. Evaluation of physicians' professional performance: an iterative development and validation study of multisource feedback instruments. *BMC Health Serv Res*. 2012;12:80.
- Wright C, Richards SH, Hill JJ, et al. Multisource feedback in evaluating the performance of doctors: the example of the UK General Medical Council patient and colleague questionnaires. *Acad Med*. 2012;87(12):1668-1678.
- Lockyer J. Multisource feedback: can it meet criteria for good assessment? *J Contin Educ Health Prof*. 2013;33(2):89-98.
- General Medical Council. Colleague and patient feedback for revalidation. http://www.gmc-uk.org/doctors/revalidation/colleague_patient_feedback.asp. Published 2018. Accessed Jan 27, 2018.
- 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(3):187-193.
- Campbell J, Wright C, Primary Care Research Group, Peninsula College of Medicine & Dentistry. GMC Multi-Source Feedback Questionnaires. Interpreting and handling multisource feedback results: Guidance for appraisers. General Medical Council Web site. https://www.gmc-uk.org/Information_for_appraisers.pdf_48212170.pdf. Published Feb 1, 2012. Accessed Jan 27, 2018.
- Elliott MN, Edwards C, Angeles J, Hambarsoomians K, Hays RD. Patterns of unit and item nonresponse in the CAHPS Hospital Survey. *Health Serv Res*. 2005;40(6 Pt 2):2096-2119.
- Ipsos MORI. *GP Patient Survey - Technical Annex: 2016-17 Annual Report*. London, UK: NHS England; 2017. <https://gp-patient.co.uk/downloads/archive/2017/GPPS%202017%20Technical%20Annex%20PUBLIC.pdf>. Pulished Jul 6, 2017. Accessed Jan 27, 2018.
- Klein DJ, Elliott MN, Haviland AM, et al. Understanding nonresponse to the 2007 Medicare CAHPS survey. *Gerontologist*. 2011;51(6):843-855.
- Martino SC, Weinick RM, Kanouse DE, et al. Reporting CAHPS and HEDIS data by race/ethnicity for Medicare beneficiaries. *Health Serv Res*. 2013;48(2 Pt 1):417-434.
- O'Malley AJ, Zaslavsky AM, Elliott MN, Zaborski L, Cleary PD. Case-mix adjustment of the CAHPS Hospital Survey. *Health Serv Res*. 2005;40(6 Pt 2):2162-2181.
- Asprey A, Campbell JL, Newbould J, et al. Challenges to the credibility of patient feedback in primary healthcare settings: a qualitative study. *Br J Gen Pract*. 2013;63(608):e200-e208.

30. Boiko O, Campbell JL, Elmore N, Davey AF, Roland M, Burt J. The role of patient experience surveys in quality assurance and improvement: a focus group study in English general practice. *Health Expect*. 2015;18(6):1982-1994.
31. Farrington C, Burt J, Boiko O, et al. Doctors' engagements with patient experience surveys in primary and secondary care: a qualitative study. *Health Expect*. 2017;20(3):385-394.
32. Kenny DA, Veldhuijzen W, Weijden Tv, et al. Interpersonal perception in the context of doctor-patient relationships: a dyadic analysis of doctor-patient communication. *Soc Sci Med*. 2010;70(5):763-768.
33. Roberts MJ, Campbell JL, Richards SH, Wright C. Self-other agreement in multisource feedback: the influence of doctor and rater group characteristics. *J Contin Educ Health Prof*. 2013;33(1):14-23.
34. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296(9):1094-1102.
35. Eva KW, Regehr G. "I'll never play professional football" and other fallacies of self-assessment. *J Contin Educ Health Prof*. 2008;28(1):14-19.
36. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med*. 1991;66(12):762-769.
37. Campbell J, Smith P, Nissen S, Bower P, Elliott M, Roland M. The GP Patient Survey for use in primary care in the National Health Service in the UK—development and psychometric characteristics. *BMC Fam Pract*. 2009;10:57.
38. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol*. 1999;77(6):1121-1134.
39. Carter TJ, Dunning D. Faulty self-assessment: why evaluating one's own competence is an intrinsically difficult task. *Soc Personal Psychol Compass*. 2008;2(1):346-360.
40. Roberts MJ, Campbell JL, Abel GA, et al. Understanding high and low patient experience scores in primary care: analysis of patients' survey data for general practices and individual doctors. *BMJ*. 2014;349:g6034.
41. Burt J, Campbell J, Abel G, et al. Improving patient experience in primary care: a multimethod programme of research on the measurement and improvement of patient experience. *Programme Grants for Applied Research*. 2017;5:1-452.
42. Burt J, Abel G, Elmore NL, et al. Rating communication in GP consultations: the association between ratings made by patients and trained clinical raters. *Med Care Res Rev*. 2018;75(2):201-218.
43. Lyratzopoulos G, Elliott M, Barbiere JM, et al. Understanding ethnic and other socio-demographic differences in patient experience of primary care: evidence from the English General Practice Patient Survey. *BMJ Qual Saf*. 2012;21(1):21-29.
44. Burt J, Abel G, Elmore N, et al. Assessing communication quality of consultations in primary care: initial reliability of the Global Consultation Rating Scale, based on the Calgary-Cambridge Guide to the Medical Interview. *BMJ Open*. 2014;4(3):e004339.
45. Ferguson J, Wakeling J, Bowie P. Factors influencing the effectiveness of multisource feedback in improving the professional practice of medical doctors: a systematic review. *BMC Med Educ*. 2014;14:76.
46. Al Ansari A, Donnon T, Al Khalifa K, Darwish A, Violato C. The construct and criterion validity of the multi-source feedback process to assess physician performance: a meta-analysis. *Adv Med Educ Pract*. 2014;5:39-51.
47. Burt J, Newbould J, Abel G, et al. Investigating the meaning of 'good' or 'very good' patient evaluations of care in English general practice: a mixed methods study. *BMJ Open*. 2017;7(3):e014718.
48. Lyratzopoulos G, Elliott MN, Barbiere JM, et al. How can health care organizations be reliably compared?: Lessons from a national survey of patient experience. *Med Care*. 2011;49(8):724-733.
49. Roland M, Elliott M, Lyratzopoulos G, 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.