

# Ratings of Physician Communication by Real and Standardized Patients

Kevin Fiscella, MD, MPH<sup>1</sup>

Peter Franks, MD<sup>2</sup>

Malathi Srinivasan, MD<sup>2</sup>

Richard L. Kravitz, MD, MSPH<sup>2</sup>

Ronald Epstein, MD<sup>1</sup>

<sup>1</sup>University of Rochester School of Medicine, Rochester, NY

<sup>2</sup>University of California Davis School of Medicine, Davis, Calif

---

## ABSTRACT

**PURPOSE** Patient ratings of physician's patient-centered communication are used by various specialty credentialing organizations and managed care organizations as a measure of physician communication skills. We wanted to compare ratings by real patients with ratings by standardized patients of physician communication.

**METHODS** We assessed physician communication using a modified version of the Health Care Climate Questionnaire (HCCQ) among a sample of 100 community physicians. The HCCQ measures physician autonomy support, a key dimension in patient-centered communication. For each physician, the questionnaire was completed by roughly 49 real patients and 2 unannounced standardized patients. Standardized patients portrayed 2 roles: gastroesophageal disorder reflux symptoms and poorly characterized chest pain with multiple unexplained symptoms. We compared the distribution, reliability, and physician rank derived from using real and standardized patients after adjusting for patient, physician, and standardized patient effects.

**RESULTS** There were real and standardized patient ratings for 96 of the 100 physicians. Compared with standardized patient scores, real-patient-derived HCCQ scores were higher (mean 22.0 vs 17.2), standard deviations were lower (3.1 vs 4.9), and ranges were similar (both 5-25). Calculated real patient reliability, given 49 ratings per physician, was 0.78 (95% confidence interval [CI], 0.71-0.84) compared with the standardized patient reliability of 0.57 (95% CI, 0.39-0.73), given 2 ratings per physician. Spearman rank correlation between mean real patient and standardized patient scores was positive but small to moderate in magnitude, 0.28.

**CONCLUSION** Real patient and standardized patient ratings of physician communication style differ substantially and appear to provide different information about physicians' communication style.

*Ann Fam Med* 2007;5:151-158. DOI: 10.1370/afm.643.

---

## INTRODUCTION

Patient-centered care is a core dimension of health care quality, and strong communication skills are required to ensure such care.<sup>1</sup> Yet, policy makers and researchers do not know how best to measure these physician communication skills.<sup>2</sup> Because patient ratings of physician communication from surveys have face validity and are readily available, they are widely used by such organizations as the American Board of Internal Medicine and various managed care organizations.<sup>3</sup> These ratings may eventually be used for specialty recredentialing and to determine physician bonus payments.

Patient ratings of physicians are subject to a number of limitations, however. These limitations include potential biases from survey nonresponse, patient selection (and deselection) of the physician, patient accommodation to the physician's style with time,<sup>4</sup> length of relationship, different patient complaints, and ceiling effects.<sup>5</sup> Patients also respond in a global manner when rating their physicians, with high correlations among scales measuring

*Conflicts of interest: none reported*

### CORRESPONDING AUTHOR

Kevin Fiscella, MD, MPH  
1381 South Ave  
Rochester, NY 14620  
Kevin\_Fiscella@urmc.rochester.edu

different behaviors and constructs.<sup>4,6</sup> Patient rating of physician communication, but not technical quality, is a key predictor of patient global satisfaction.<sup>7</sup>

Standardized patients—persons trained to portray a specific patient case in a standardized fashion<sup>8</sup>—represent a potentially more objective means for assessment of physician communication. Standardized patients can be trained to rate physician-patient communications skills reliably<sup>9,10</sup> and are used by medical schools and the US Medical Licensure Examination to assess competence.<sup>8,11,12</sup> Unannounced standardized patients have also been used in research to rate community physicians' history taking, physical examination, medical decision making, and communication, as well as health services utilization and quality of care.<sup>13-18</sup> They have been regarded by some as the reference standard for assessing physician performance,<sup>19</sup> but their use in assessing communication is still evolving. In addition to assessing specific behaviors, standardized patients can also provide their subjective impression of physicians. Their training and exposure to multiple physicians allows standardized patients to act as connoisseurs of clinical care.<sup>20</sup>

To better understand the strengths and limitations of real patient and standardized patient assessment techniques, we sought to compare standardized patient ratings with real patient ratings of physicians' patient-centered communication. Specifically, we compared the distribution of scores on identical scales completed by real patients and standardized patients, assessed the correlation between real and standardized patient ratings, and contrasted differences in rank order using these 2 assessment methods. We used an adaptation of the Health Care Climate Questionnaire (HCCQ) to assess communication. The HCCQ is designed to assess a key component of physician communication, that is, the extent to which the patient perceives the physician as supporting patient autonomy in decision making.<sup>21</sup>

## METHODS

We compared perceptions of primary care physicians' patient-centered communication from 2 sources: (1) standardized patient ratings, 2 encounters per physician, and (2) real patient ratings, completed by about 49 patients of each physician. Real patients also provided information regarding their sociodemographic characteristics, health status, and duration of their relationship with their physician. Full details regarding the study are provided elsewhere.<sup>4,17,22</sup>

### Physician Sample

In 1999, we identified 594 primary care physicians (internists and family physicians) in active clinical

practice within 45 minutes' drive of Rochester, NY, who belonged to a large managed care organization (MCO) serving the 8-county Rochester region (population 1.1 million). To achieve stable measures of performance indicators, only the 506 physicians who had more than 100 MCO patients were eligible; thus, enrolled physicians had larger practices than nonenrolled physicians. A maximum of 2 physicians per practice were recruited to avoid clustering effects and to minimize physician detection of standardized patients. In all, 297 physicians were personally contacted by 12 physician-recruiters in random order until a total of 100 physicians were recruited. Of the 297 physicians identified for recruitment, 109 (37%) refused to participate, and 14 were ineligible. Analyses of MCO claims data showed that the patient and utilization characteristics among participating and nonparticipating practices were similar.

Physicians gave informed consent to participate in a study of "patient care and outcomes." They agreed to have 2 unannounced, covertly audio-recorded standardized patient visits within the subsequent 12 months (2000-2001). Physicians were reimbursed \$100 for each standardized patient visit (including administration of waiting room patient questionnaires); \$100 was also provided to the office staff for their assistance. The study received institutional review board approval from the University of Rochester and local hospitals.

### Standardized Patient Roles

We developed standardized patient roles to provide physicians with identical clinical scenarios and to avoid potential rating biases with real patients. A physician community advisory panel suggested that using more than 2 standardized patients per physician would likely compromise physician recruitment. Thus, during the 2000-2001 period, each participating physician had visits from 2 standardized patients, portraying 1 of 2 roles. The 2 roles were selected to contrast physicians' responses to real patients who had straightforward symptoms with those who reported ambiguous or medically unexplained symptoms. The straightforward role involved a patient with symptoms typical of gastroesophageal reflux disorder (GERD) or heartburn, complaining of nocturnal chest pain exacerbated by food and partially relieved by antacids, and with minimal emotional distress. The ambiguous symptoms role involved a patient with symptoms characteristic of multisomatoform disorder: multiple symptoms, poorly characterized chest pain, and moderate emotional distress.<sup>23</sup> During pilot testing, the roles were calibrated to avoid prompting referral to an emergency department or administration of medications while in the physician's office. Five middle-aged, nonobese, white

standardized patients (2 men, 3 women) were used in the study. Each physician saw 1 male and 1 female standardized patient, and 1 of each role (2 female standardized patients were needed for the GERD role, as 1 standardized patient could not complete all visits in the study; each standardized patient made at least 25 visits). The first standardized patient visit was randomized by illness condition and sex.

### Standardized Patient and Real Patient Survey Data

We use the HCCQ, based on self-determination theory,<sup>21</sup> to assess patient-centered communication. The HCCQ assesses physician support for patient autonomy. Autonomy support is associated with improved patient adherence, outcomes, and satisfaction, and is a critical element of patient-centered communication.<sup>24-27</sup> Items in the scale include judgments about whether the physician provided patients with options about their health, conveyed confidence to them in their ability to make changes important for their health, and tried to understand their perspectives before suggesting medical or behavioral changes. Items also assess whether the patient felt understood by the physician, and whether the physician encouraged the patient to ask questions. The modified 5-item version of the HCCQ uses a 5-point Likert scale (1 = strongly disagree and 5 = strongly agree) to rate physician behavior, with a score range of 5 (low support) to 25 (high support).

After each encounter, the standardized patient completed the HCCQ. Cronbach's  $\alpha$  for the scale, as reported by the standardized patients, was 0.92.

About 50 consecutive real patients aged 18 to 65 years were approached by a research assistant in each of the primary care physician's waiting room to complete a questionnaire before their office visit. The questionnaire included the HCCQ and questions on demographics (age, sex, race/ethnicity, and years of schooling), health status (a checklist of chronic disease conditions), the Medical Outcomes Study Short Form 12-item survey (SF-12),<sup>28</sup> and duration (years) of the patient's relationship with the physician. Cronbach's  $\alpha$  for the HCCQ scale, as reported by real patients, was also 0.92.

About halfway through our project, a new instrument became available for patients to assess their perception of the physician's patient-centeredness during office visits. This instrument, called the Patient Perception of Patient-Centeredness (PPPC) was developed by Stewart et al.<sup>29</sup> It comprises 14 Likert-scale items, each with a 4-point scale. Standardized patients and real patients completed the PPPC measure immediately after their visit with participating physicians. The

PPPC measure was scaled from a minimum of 14 (least patient-centered) to 56 (most patient-centered). The Cronbach's  $\alpha$  for the 14-item postvisit PPPC scale in patients was 0.90. We examined the physician  $\rho$  statistic for both real and standardized patients for the PPPC, and also compared real and standardized patient scores for the HCCQ and PPPC.

### Statistical Analyses

Data were analyzed using Stata (Version 9.2, StataCorp, College Station, Tex), and analyses were adjusted for the nesting of standardized and real patient observations by physician. We conducted 2 sets of random intercept, mixed-effects regression analyses with the HCCQ scores as dependent variables and the physicians as independent random effects. In the standardized patient analyses, the other independent variables included role (GERD vs ambiguous chest pain) and a dummy variable for each standardized patient (to adjust for systematic differences in standardized patient scoring unrelated to the physician effects). In the real patient analyses, the other independent variables included factors that might bias patients' assessment of physician interpersonal style (demographics, health status, and relationship length). We did not include physician, physician practice, or health system factors, because these variables were similarly experienced by real and standardized patients. Assessment of the physician random effects (or variance components) allowed (1) calculation of the interrater reliability of the standardized patients' and real patients' assessment of HCCQ scores, and (2) assessment of the physicians' predicted effects on HCCQ scores. We report the reliability based on the proportion of total variance contributed by the physician variance component ( $\rho$ , an intraclass correlation coefficient).<sup>30</sup> In turn, the predicted reliability, a function of the number observations, is derived from the intraclass correlation coefficient using the Spearman-Brown prophecy formula:

$$\rho_{xx'} = \frac{n \times \rho_{xx}}{1 + \rho_{xx}(n - 1)}$$

Here,  $\rho_{xx'}$  is predicted (or desired) reliability,  $n$  is the number of observations, and  $\rho_{xx}$  is observed intraclass correlation coefficient (observed reliability) from our study. The predicted effects of the physicians on the patient-centered communication score, as derived using a random effects model, is known as the best linear unbiased prediction. Key advantages of the random effects model compared with a fixed dummy variable approach to assessing physician performance include allowing inferences based on the total possible universe of samples (beyond the predicted effects just based on the specific physicians and raters sampled),

**Table 1. Characteristics of Physicians in Sample (N = 100)**

Characteristic	% With Characteristic	Characteristic	% With Characteristic
Sex		Solo practitioner	
Female	23	Yes	24
Male	77	No	76
Family physician		Rural practice	
Yes	47	Yes	32
No	53	No	68

Note: results are unadjusted.

**Table 2. Characteristics of Real Patients in the Sample**

Characteristic	Value
Total, No. (%)	4,746 (100)
Age, years, mean (SD)	44.9 (12.1)
Sex, No. (%)	
Female	2,955 (62.3)
Male	1,750 (36.9)
Race, No. (%)	
African American	499 (10.5)
Hispanic	109 (2.3)
Other	110 (2.3)
White	3,994 (84.2)
Education, No. (%)	
<12 years	337 (7.1)
12th grade	1,370 (28.9)
1-3 years college	1,470 (31.2)
4 years college	828 (17.4)
Graduate school	700 (14.7)
Length of patient-doctor relationship, No. (%)	
<1 years	360 (7.6)
1-3 years	1,035 (21.8)
3-5 years	814 (17.2)
>5 years	2,525 (53.2)
Number of medical conditions, mean (SD)	2.3 (1.5)
MCS-12	48.8 (10.5)
PCS-12	46.0 (11.1)
HCCQ	8.0 (3.1)

\*Numbers differ slightly due to missing value.

MCS-12 = medical component summary of the SF-12 Health Survey; PCS-12 = physical component summary of the SF-12 Health Survey; HCCQ = Health Care Climate Questionnaire.

and adjustment for the variable uncertainty in the predicted effects caused by differing amounts of information about each physician (for example, the number of raters).<sup>30</sup> We explore the shared variance between real and standardized patient ratings of physicians, based on their Pearson product moment correlations, and the differences in how they ranked physicians using Spearman rank order correlations. Ranking was based on standardized-patient-derived postvisit HCCQ scores and real-patient-derived previsit scores.

## RESULTS

Of the 100 physicians, 93 completed both standardized patient visits, and 7 completed only 1 visit because of retirement or relocation; no other physician withdrew from the study. The physicians had a mean age of 45 years and were predominantly male; most were in a group practice and in a nonrural location (Table 1). A total of 4,746 patients (96% of those approached) were enrolled and completed the previsit questionnaire. In all, questionnaires were completed by patients for 96 of the 100 physicians. Enrolled patients were modally female, white, had at least some college education, and at least a 5-year relationship with their physician (Table 2). Complete HCCQ and demographic data were available for approximately 49 patients per physician. For the patient-centered communication postvisit subset, we obtained data from 1,730 patients of 57 doctors, with the same eligibility criteria and methods as the previsit questionnaires.

Real patient HCCQ scores were skewed toward favorable physician ratings. The mean patient HCCQ score was 22.0 (of a possible 25), with a standard deviation 3.1, and a range 5 to 25. The unannounced standardized patient scores were more normally distributed. Their mean HCCQ score was lower (17.25), and the standard deviation was greater (4.9), but with a similar range (5-25).

Adjusting for real patient demographics, health status, and length of relationship, the physician  $\rho$  for autonomy support based on real patient ratings was 0.07 (95% confidence interval [CI], 0.05-0.108;  $P < .001$ ). Using the Spearman-Brown prophecy formula, based on an average of 49 real patients per physician, the patient-based physician  $\rho$  translates into a reliability of 0.78 (95% CI, 0.71-0.84). To reach a desired reliability of 0.9, ratings from 120 patients (95% CI, 75-180) would be required. After adjustment for standardized patient and case, the physician  $\rho$  for autonomy support based on standardized patient ratings was 0.40 (95% CI, 0.24-0.58;  $P < .001$ ). Using the Spearman-Brown prophecy formula, based on 2 observations per physician, the standardized-patient-based physician  $\rho$  translates into a reliability of 0.57 (95% CI, 0.39-0.73). To reach a desired reliability of 0.9, ratings from 14 standardized patients (95% CI, 7-29) would be required.

The Pearson product moment correlation between the real-patient- and standardized-patient-derived, best, linear, unbiased predicted physician scores was 0.30, which, when squared, yields the shared variance, 9%. The maximum possible correlation, given the imperfect standardized and real patient reliability,

is the square root of the product of their reliabilities (or 0.67). Thus, the adjusted correlation, given perfect reliability, would be 0.47. In other words, the maximum shared variance between the 2 sets of scores is 22%. The relationship between the 2 sets of scores is shown in Figure 1. As can be seen, standardized patient and real patient rankings were quite different, with a Spearman rank order correlation of 0.28.

For the 57 physicians on whom postvisit PPPC ratings were obtained, the mean PPPC score for real patients was higher, 51.5 (SD 5.7; range, 19-56), than that for standardized patients, 38.2 (SD 9.2; range, 19-55). After adjustment for standardized patient and case, the physician  $\rho$  for PPPC based on standardized patient ratings was 0.40 (95% CI, 0.21-0.63;  $P < .001$ ). Adjusting for patient demographics, health status and length of relationship, the physician  $\rho$  for PPPC based on real patient ratings was 0.06 (95% CI, 0.03-0.09;  $P < .001$ ). The correlations among the physician-level mean HCCQ and PPPC scores by real and standardized patients are shown in Table 3.

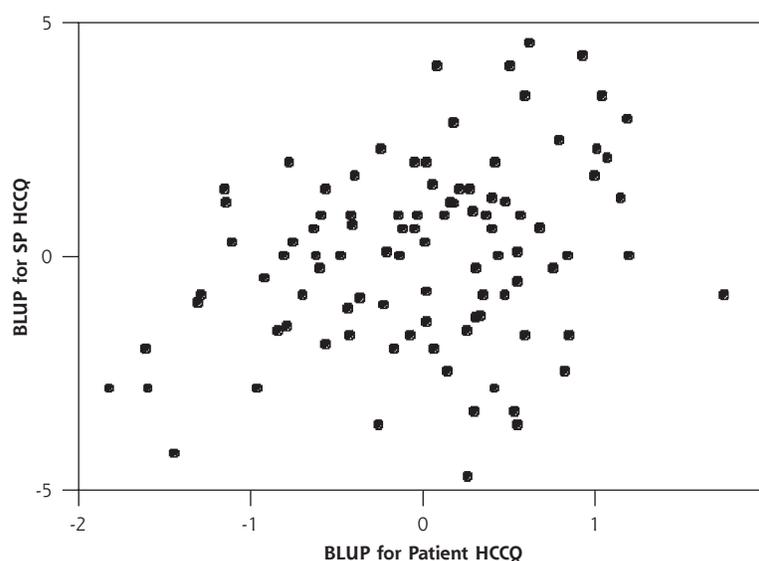
## DISCUSSION

Although there is no external standard by which to judge the relative validity of either real or standardized patient ratings, unannounced standardized patient ratings of physician communication skills show superior psychometric properties. In contrast to real patient ratings, standardized patient ratings were more normally distributed and showed greater variation. The proportion of total variance contributed by the physician variance component was much greater for standardized than real patient ratings. As a result, a single postvisit standardized patient rating is far more reliable than a single real patient self-report, either before or after a clinic visit. The low correlations and differences in outliers show that real and standardized patient rating ratings yield quite different rankings of physicians for autonomy support. These correlations, using a sample of community physicians, are somewhat lower than those observed in a similarly

designed study of 26 Canadian resident physicians.<sup>31</sup> In that study, an adjusted correlation of 0.51 was observed between standardized patient and real patient ratings using the American Board of Internal Medicine satisfaction scale.

The HCCQ data were obtained after the standardized patient encounter and before the real patient encounter, whereas the PPPC data were obtained after the encounter for both standardized and real patients. Despite these methods differences, the results for the HCCQ and PPPC closely parallel each other. Thus, for the standardized patients, mean PPPC scores were lower and standard deviations were higher than observed for real patients. Further, for both real and standardized patient evaluations, the physician  $\rho$  for the PPPC scores was quite similar to the physician  $\rho$  obtained for the HCCQ; and for both PPPC and HCCQ, the physician  $\rho$  was much higher for standardized patient evaluations than for real patient evaluations. The high correlations between the standardized

**Figure 1. Relationship between adjusted standardized patient and real patient ratings of physicians.**



BLUP = best linear unbiased prediction; SP = standardized patient; HCCQ = Health Care Climate Questionnaire.

**Table 3. Correlations Among the Physician-Level Mean HCCQ and PPPC Scores Obtained From Real Patients and Standardized Patients**

Measure	Real Patient HCCQ (Previsit)	Real Patient PPPC (Postvisit)	SP HCCQ (Postvisit)
Real patient PPPC (postvisit)	0.74	—	—
SP HCCQ (postvisit)	0.33	0.32	—
SP PPPC (postvisit)	0.33	0.26	0.89

HCCQ = Health Care Climate Questionnaire; PPPC = Patient Perception of Patient Centeredness questionnaire; SP = standardized patient.

patient HCCQ and PPPC scales (0.89) suggest that both scales measure a similar underlying construct. This observation is consistent with previous work<sup>6</sup> suggesting that when assessing physicians along nominally different dimensions, patients tend to make a global, or halo, assessment; thus, at the physician level, substantive distinctions among the constructs may not be critical. Despite data being obtained before (HCCQ) and after (PPPC) the encounter, the correlations between the 2 patient-scored scales were also high (0.74).

For both scales, the standardized and real patient correlations were modest (between 0.26 and 0.33). The low correlation in physician ranking based on real and standardized patient ratings suggest caution in using only real patient ratings for pay-for-performance or recertification. Different physicians would be rewarded for their interpersonal skills depending on which measurement approach is used. These findings, combined with those from studies suggesting that associations between patient ratings and patient report of health status and other outcomes may be confounded by patients perceptual styles<sup>4</sup> (eg, negative patient influence may influence both physician ratings and self-reported health), suggest the need for alternative ways of assessing physician communication.

Our findings are consistent with our original hypothesis that trained standardized patients provide a more objective (though much narrower) assessment of physician communication skills. The differences in ratings may be due to the biases that each group brings to physician assessment. For instance, patients often self-select physicians with whom they develop a longitudinal relationship. Dissatisfied patients, particularly patients who are dissatisfied with the quality of their interpersonal relationship with their physician, are more likely to change physicians.<sup>32,33</sup> Thus, remaining patients represent a selected group in part based on their level of satisfaction with their physicians. Patients may also accommodate themselves to physicians' interpersonal styles with time. Halo effects may further bias patient assessment; patients who like their physicians are more likely to provide favorable ratings regardless of the aspect of physician performance being assessed. The combination of continued exposure to their physicians' interpersonal styles, limited exposure to alternative physician styles, and cognitive dissonance (eg, "Would I continue to see a physician who treats me badly?"), all tend to mitigate differences between physicians based on patient ratings.

In contrast with real patients, standardized patients are trained evaluators who have the opportunity to visit multiple physicians. As a result, they may become connoisseurs of communication skills. They are able to assess a physician's communication skills based on their

interactions with many other physicians over a relatively short period and under highly controlled conditions. Their perceptions are relatively unencumbered by prior experience with the physician, sense of indebtedness, or personal investment. They have no agendas or specific expectations for the visit that might affect their ratings.

Others have suggested that standardized patients are a feasible and practical means for assessing physician performance,<sup>13,19,31,34,35</sup> although these suggestions are based on standardized patient assessment of specific behaviors rather than global impressions. In either case, however, their use requires considerable efforts to avoid detection by physicians, including surreptitious collaboration with office staff. Such complicity might prove problematic for pay-for-performance purposes, as office staff might tip the physician to the identity of the unannounced standardized patient. Some data, however, suggest that physician detection of standardized patients has little impact on physician performance.<sup>36</sup>

Although standardized patients have been viewed as the reference standard for assessing physician management of specific problems,<sup>13</sup> their ratings are not without limitations. Standardized patients typically rate physicians based on a single first visit devoid of previous context. They are thus more useful for assessing new-patient visits than for assessing chronic care, a fundamental aspect of primary care. Because they have only one opportunity to interact with the physician, standardized patients respond to an important first impression. That first impression may not capture some essential elements of physician care over time. For instance, standardized patients would not be able to assess how physicians would tailor interventions based on close knowledge of patient preferences or family, whether the physician would be supportive during times of emotional or personal distress, or how well they advocate for their patients in complex health systems.<sup>37</sup> Ratings derived from white, middle-class standardized patients might differ from those from standardized patients who are members of racial or ethnic minority groups. Finally, there are few data that relate standardized patient evaluations to meaningful clinical outcomes. Despite these limitations, we believe that standardized patients represent a reasonably objective means of assessing specific physician communication skills exhibited under tightly controlled conditions.

Our findings should be considered in the context of the study limitations. First, real patient ratings were elicited previsit, whereas standardized patients completed their rating immediately following their visit. Previsit ratings reflect the cumulative experience of the patient with the physician and provide a general assessment of the physician. Postvisit assessments provide a visit-spe-

cific assessment based on more detailed recall of patient-physician interactions. These two types of ratings are only modestly correlated.<sup>38</sup> Even though it is possible that higher correlations would have been observed had we surveyed all patients postvisit, our subanalysis using PPPC data suggests otherwise. Moreover, postvisit surveys are more challenging to administer; many instruments are discarded, resulting in potential response bias. As a result, patient ratings, such as the Consumer Assessment of Health Plan Survey, are administered by telephone or mail without regard to the last visit.<sup>39</sup> Thus, real patient ratings observed in our study more closely reflect general ratings of physician communication that are currently being used in practice.

Second, standardized patient ratings were based on only 2 different cases. Whether these findings generalize to the range of problems treated by primary care physicians cannot be determined from these data. Third, we obtained a very high response to our patient survey. Response rates to most patient surveys are considerably lower and thus subject to greater response bias.<sup>40,41</sup> Last, the physician sample required voluntary enrollment in a study on patient-physician communication. Communication styles of participating physicians may have differed from those who did not participate. It is possible that greater variability and perhaps lower ratings would be seen among a less-selected sample.

For these reasons, it is not possible from these data to determine for certain the reasons for the low correlations between real and standardized patient ratings. It is not clear whether they are primarily due to differences in context (eg, longitudinal relationship vs one-time visit), methods of administration for the HCCQ scale, or real differences in how well these 2 methods assess a key aspect of physician-patient communication, autonomy supportiveness. Further research using comparable types of visits (eg, new patients) and methods (postvisit assessment by real patients) is needed to clarify these important questions.

In conclusion, we found that unannounced standardized patient ratings have better psychometric properties than real patients' ratings of physicians' communication behaviors, and that real and standardized patient ratings correlated poorly. These findings raise questions regarding the validity of patient ratings alone for recertifying physicians or for pay-for-performance incentive programs and the potential usefulness of multiple methods for studying physician interactional behavior.<sup>2</sup>

**To read or post commentaries in response to this article, see it online at <http://www.annfam.org/cgi/content/full/5/2/151>.**

**Key words:** Physician-patient relations; clinical competence; reproducibility of results; patient simulation

Submitted May 4, 2006; submitted, revised, August 3, 2006; accepted August 14, 2006.

**Funding support:** Supported through funding by the Agency for Healthcare Quality and Research (1 R01 10601-A1) and National Institute of Mental Health (5 R01 MH064683-03).

## References

1. Institutes of Medicine. *Envisioning the National Health Care Quality Report*. Washington, DC: National Academy Press; 2001.
2. Epstein RM, Franks P, Fiscella K, et al. Measuring patient-centered communication in patient-physician consultations: theoretical and practical issues. *Soc Sci Med*. 2005;61(7):1516-1528.
3. Marshall GN, Morales LS, Elliott M, Spritzer K, Hays RD. Confirmatory factor analysis of the Consumer Assessment of Health Plans Study (CAHPS) 1.0 Core Survey. *Psychol Assess*. 2001;13(2):216-229.
4. Franks P, Fiscella K, Shields CG, et al. Are patients' ratings of their physicians related to health outcomes? *Ann Fam Med*. 2005;3(3):229-234.
5. Sitzia J. How valid and reliable are patient satisfaction data? An analysis of 195 studies. *Int J Qual Health Care*. 1999;11(4):319-328.
6. Franks P, Jerant AF, Fiscella K, et al. Studying physician effects on patient outcomes: physician interactional style and performance on quality of care indicators. *Soc Sci Med*. 2006;62(2):422-432.
7. Chang JT, Hays RD, Shekelle PG, et al. Patients' global ratings of their health care are not associated with the technical quality of their care. *Ann Intern Med*. 2006;144(9):665-672.
8. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002;287(2):226-235.
9. McLaughlin K, Gregor L, Jones A, Coderre S. Can standardized patients replace physicians as OSCE examiners? *BMC Med Educ*. 2006;6:12.
10. Guiton G, Hodgson CS, Delandshere G, Wilkerson L. Communication skills in standardized-patient assessment of final-year medical students: a psychometric study. *Adv Health Sci Educ Theory Pract*. 2004;9(3):179-187.
11. Chessman AW, Blue AV, Gilbert GE, Carey M, Mainous AG, 3rd. Assessing students' communication and interpersonal skills across evaluation settings. *Fam Med*. 2003;35(9):643-648.
12. Colliver JA, Swartz MH. Assessing clinical performance with standardized patients. *JAMA*. 1997;278(9):790-791.
13. Glassman PA, Luck J, O'Gara EM, Peabody JW. Using standardized patients to measure quality: evidence from the literature and a prospective study. *Jt Comm J Qual Improv*. 2000;26(11):644-653.
14. Ramsey PG, Curtis JR, Paauw DS, Carline JD, Wenrich MD. History-taking and preventive medicine skills among primary care physicians: an assessment using standardized patients. *Am J Med*. 1998;104(2):152-158.
15. Carney PA, Ward DH. Using unannounced standardized patients to assess the HIV preventive practices of family nurse practitioners and family physicians. *Nurse Pract*. 1998;23(2):56-58, 63, 67-58 passim.
16. Gallagher TH, Lo B, Chesney M, Christensen K. How do physicians respond to patient's requests for costly, unindicated services? *J Gen Intern Med*. 1997;12(11):663-668.
17. Epstein RM, Franks P, Shields CG, et al. Patient-centered communication and diagnostic testing. *Ann Fam Med*. 2005;3(5):415-421.
18. Kravitz RL, Epstein RM, Feldman MD, et al. Influence of patients' requests for direct-to-consumer advertised antidepressants: a randomized controlled trial. *JAMA*. 2005;293(16):1995-2002.
19. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA*. 2000;283(13):1715-1722.

20. Srinivasan M, Franks P, Meredith L, et al. Connoisseurs of care? Unannounced standardized patients' ratings of physicians. *Med Care*. 2006;44(12):1092-1098.
21. Williams GC, McGregor HA, Zeldman A, Freedman ZR, Deci EL. Testing a self-determination theory process model for promoting glycemic control through diabetes self-management. *Health Psychol*. 2004;23(1):58-66.
22. Fiscella K, Meldrum S, Franks P, et al. Patient trust: is it related to patient-centered behavior of primary care physicians? *Med Care*. 2004;42(11):1049-1055.
23. Kroenke K, Spitzer RL, deGruy FV, 3rd, et al. Multisomatoform disorder. An alternative to undifferentiated somatoform disorder for the somatizing patient in primary care. *Arch Gen Psychiatry*. 1997;54(4):352-358.
24. Williams GC, Grow VM, Freedman ZR, Ryan RM, Deci EL. Motivational predictors of weight loss and weight-loss maintenance. *J Pers Soc Psychol*. 1996;70(1):115-126.
25. Williams GC, Freedman ZR, Deci EL. Supporting autonomy to motivate patients with diabetes for glucose control. *Diabetes Care*. 1998;21(10):1644-1651.
26. Greenfield S, Kaplan SH, Ware JE, Jr., Yano EM, Frank HJ. Patients' participation in medical care: effects on blood sugar control and quality of life in diabetes. *J Gen Intern Med*. 1988;3(5):448-457.
27. Cecil DW, Killeen I. Control, compliance, and satisfaction in the family practice encounter. *Fam Med*. 1997;29(9):653-657.
28. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220-233.
29. Stewart M, Brown JB, Donner A, et al. The impact of patient-centered care on outcomes. *J Fam Pract*. 2000;49(9):796-804.
30. Snijders T, Bosker R. *Multilevel Analysis: An Introduction to Basic and Multilevel Modeling*. London: Sage Publications; 1999.
31. Tamblyn R, Abramowicz M, Schnarch B, et al. Can standardized patients predict real-patient satisfaction with the doctor-patient relationship? *Teach Learn Med*. 1994;6:36-44.
32. Federman AD, Cook EF, Phillips RS, et al. Intention to discontinue care among primary care patients: influence of physician behavior and process of care. *J Gen Intern Med*. 2001;16(10):668-674.
33. Safran DG, Montgomery JE, Chang H, Murphy J, Rogers WH. Switching doctors: predictors of voluntary disenrollment from a primary physician's practice. *J Fam Pract*. 2001;50(2):130-136.
34. Tamblyn RM. Use of standardized patients in the assessment of medical practice. *CMAJ*. 1998;158(2):205-207.
35. Carney PA, Dietrich AJ, Freeman DH, Jr., Mott LA. A standardized-patient assessment of a continuing medical education program to improve physicians' cancer-control clinical skills. *Acad Med*. 1995;70(1):52-58.
36. Franz C, Epstein R, Miller K, et al. Caught in the act? Prevalence, predictors, and consequences of physician detection of unannounced standardized patients. *Health Serv Res*. 2006;41(6):2290-3202.
37. Kravitz RL, Franks P, Feldman M, et al. What drives referral from primary care physicians to mental health specialists? A randomized trial using actors portraying depressive symptoms. *J Gen Intern Med*. 2006;21(6):584-589.
38. Marshall BS, Long MJ, Voss J, Demma K, Skerl KP. Case management of the elderly in a health maintenance organization: the implications for program administration under managed care. *J Healthc Manag*. 1999;44(3):477-491; discussion 492-473.
39. Fowler FJ, Jr., Gallagher PM, Stringfellow VL, et al. Using telephone interviews to reduce nonresponse bias to mail surveys of health plan members. *Med Care*. 2002;40(3):190-200.
40. Zaslavsky AM, Zaboriski LB, Cleary PD. Factors affecting response rates to the Consumer Assessment of Health Plans Study survey. *Med Care*. 2002;40(6):485-499.
41. Weech-Maldonado R, Morales LS, Elliott M, et al. Race/ethnicity, language, and patients' assessments of care in Medicaid managed care. *Health Serv Res*. 2003;38(3):789-808.