

UC Davis

UC Davis Previously Published Works

Title

Reliability of Physician-Level Measures of Patient Experience in Primary Care

Permalink

<https://escholarship.org/uc/item/38f2p6w9>

Journal

Journal of General Internal Medicine, 32(12)

ISSN

0884-8734

Authors

Fenton, Joshua J

Jerant, Anthony

Kravitz, Richard L

et al.

Publication Date

2017-12-01

DOI

10.1007/s11606-017-4175-y

Peer reviewed

Reliability of Physician-Level Measures of Patient Experience in Primary Care

Joshua J. Fenton, MD, MPH^{1,2,3}, Anthony Jerant, MD^{1,2}, Richard L. Kravitz, MD, MSPH^{2,4}, Klea D. Bertakis, MD, MPH^{1,2}, Daniel J. Tancredi, PhD^{2,5}, Elizabeth M. Magnan, MD, PhD^{1,2}, and Peter Franks, MD^{1,2}

¹Department of Family and Community Medicine, University of California, Davis Medical Center, Sacramento, CA, USA; ²The Center for Healthcare Policy and Research, University of California, Davis Medical Center, Sacramento, CA, USA; ³Department of Family and Community Medicine, University of California, Davis Health System, Sacramento, CA, USA; ⁴The Division of General Medicine, Department of Internal Medicine, University of California, Davis Medical Center, Sacramento, CA, USA; ⁵Department of Pediatrics, University of California, Davis Medical Center, Sacramento, CA, USA.

BACKGROUND: Patient experience measures are widely used to compare performance at the individual physician level.

OBJECTIVE: To assess the impact of unmeasured patient characteristics on visit-level patient experience measures and the sample sizes required to reliably measure patient experience at the primary care physician (PCP) level.

DESIGN: Repeated cross-sectional design.

SETTING: Academic family medicine practice in California.

PARTICIPANTS: One thousand one hundred forty-one adult patients attending 1319 visits with 56 PCPs (including 45 resident and 11 faculty physicians).

MEASUREMENTS: Post-visit patient experience surveys including patient measures used for standard adjustment as recommend by the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Consortium and additional patient characteristics used for expanded adjustment (including attitudes toward healthcare, global life satisfaction, patient personality, current symptom bother, and marital status).

RESULTS: The amount of variance in patient experience explained doubled with expanded adjustment for patient characteristics compared with standard adjustment ($R^2 = 20.0\%$ vs. 9.6% , respectively). With expanded adjustment, the amount of variance attributable to the PCP dropped from 6.1% to 3.4% and the required sample size to achieve a reliability of 0.90 in the physician-level patient experience measure increased from 138 to 255 patients per physician. After ranking of the 56 PCPs by average patient experience, 8 were reclassified into or out of the top or bottom quartiles of average experience with expanded as compared to standard adjustment [14.3% (95% CI: $7.0\text{--}25.2\%$)].

CONCLUSIONS: Widely used methods for measuring PCP-level patient experience may not account sufficiently for influential patient characteristics. If methods were

adapted to account for these characteristics, patient sample sizes for reliable between-physician comparisons may be too large for most practices to obtain.

J Gen Intern Med 32(12):1323–9

DOI: 10.1007/s11606-017-4175-y

© Society of General Internal Medicine 2017

INTRODUCTION

Patient experience is widely considered a core dimension of healthcare quality. The Affordable Care Act mandated the Centers for Medicare and Medicaid Services (CMS) to include patient experience metrics in pay-for-performance programs. Patient experience data from US hospitals are publicly reported on CMS's Hospital Compare website,¹ and ambulatory patient experience metrics determine Shared Savings Program reimbursements for Accountable Care Organizations. After ambulatory visits, patients now commonly receive a mailed survey about their recent visit experience, and some organizations provide bonus compensation to top-performing clinicians.² Some have urged broader adoption of incentives based on physician-level patient experience measures and the incorporation of these measures into licensing and recertification requirements.³

For fair comparison of individual physicians, physician-level patient experience measurements must have high reliability. Reliability refers to the reproducibility of a measurement; if physician-level patient experience measures were highly reliable, one would expect that two patients visiting the same physician would tend to rate the physician similarly. However, if the physician impact on the experience measure is weak compared to patient, health system, or other factors, the measure may vary substantially across patients, necessitating a larger patient sample to reliably measure the individual physician effect. Small sample sizes have been shown to limit the reliability of other physician-level quality measures.^{4,5}

We collected patient experience surveys within a large academic family practice to assess: (1) the extent to which patient-level factors that are unmeasured in standard Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys explain additional variance in patient experience

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11606-017-4175-y>) contains supplementary material, which is available to authorized users.

Received February 1, 2017

Revised June 22, 2017

Accepted August 24, 2017

Published online September 12, 2017

measures; (2) the comparative size of the physician effect on patient experience with and without adjustment for these additional patient characteristics; (3) required sample sizes for reliable measurement of physician-level patient experience assuming full adjustment for potentially confounding patient characteristics; (4) the impact of unmeasured confounding by patient characteristics on physician rankings based on visit-level patient experience measures.

METHODS

Design, Setting, and Subjects. The study had a repeated cross-sectional design with patient visits being the units of analyses and with visits cross-nested within patients and primary care physicians (PCPs). Three days per week from July 2015 to April 2016, a research assistant recruited patients from the waiting room of the UC Davis Family Medicine Clinic, an urban academic practice with ~27,500 annual visits with resident and attending family physicians. Presumed patients were asked to approach the assistant after visits, if interested in participation, at which point the assistant confirmed eligibility and obtained written informed consent. Patients were eligible to participate if they were aged ≥ 18 years, could read and complete an English survey, and were attending physician visits. Consenting patients completed post-visit surveys on tablet devices, although a small number preferred paper surveys. The tablet survey was administered using LimeSurvey software, which provided real-time checks for data quality and completeness. Patients were allowed to complete surveys after up to six physician visits during the study period and were compensated with \$10 gift cards for survey completion. Research assistants could not ascertain eligibility when potential participants were approached, and due to limited staffing, assistants could not feasibly approach all potentially eligible patients attending visits. Based on the annual number of clinic visits, we estimate that patients completed surveys after ~15% of adult visits on days when research assistants were present in the clinic. The UC Davis Institutional Review Board approved the study.

Patient Experience Measures. Patient experience was assessed using six items derived from the individual visit version of the CAHPS Clinician & Group Survey.⁶ Four items derived from the CAHPS Physician Communication Composite and inquired respectively about whether the PCP: (1) gave easy to understand information; (2) knew important information about the patients' medical history; (3) showed respect for what the patient had to say; (4) spent enough time with the patient. A fifth item inquired about whether the patient would recommend the PCP to family and friends, while the sixth item requested that the patient rate the doctor from 0 to 10 from worse to best possible doctor. The six items were highly correlated and loaded onto a single latent construct in factor analyses (see [Supplemental Appendix](#)). To enhance measure reliability, we created a standardized scale

in which higher numbers indicated better patient experience by averaging the z-score for each item (Cronbach's $\alpha = 0.80$).

Standard Adjustors for CAHPS Surveys. We collected sociodemographic and clinical measures that the CAHPS Consortium recommends as standard adjustors in CAHPS data analyses, including patient age, sex, race/ethnicity, education, and self-reported health status (poor, fair, good, very good, excellent).⁷ While the CAHPS survey assesses mental and emotional health with a single question, we assessed mental health using the five-item Mental Health Inventory-5, an accurate measure of both depression and anxiety (range: 0–100 from worst to best mental health).⁸

Additional Patient-Level Adjustors. We included additional patient-level measures to enable analyses of the influence of an expanded range of covariates on patient experience measures. We included marital status because it may influence perceived healthcare experience through adverse impacts on mental health or social support.⁹ We assessed whether the patient had previously had visits with the PCP because of its previously observed strong influence on patient experience in primary care.¹⁰ Because somatic symptom burden has predicted difficult patient-doctor encounters,¹¹ we included three items regarding the extent to which patients are bothered and concerned about symptoms and worried about health. Because the items were highly correlated, we created a scale from the items in which a higher score signifies greater symptom bothersomeness/worry (range: 3–15, Cronbach's $\alpha = 0.83$). By linking patient surveys to electronic medical records, we identified the visit physician and collected patient body mass index (BMI), as higher BMI has been associated with better patient experience ratings.¹²

We also assessed three durable patient-level attitudinal or dispositional factors that we theorized could affect patient perceptions of healthcare experiences. First, we assessed skepticism regarding medical care, a four-item measure that is conceptualized as a trait that predisposes patients to use less healthcare and fewer preventive services and to make less healthful lifestyle choices.¹³ Second, we assessed global life satisfaction using the five-item Satisfaction with Life Scale, a validated measure of subjective well-being with high temporal reliability.¹⁴ Third, we assessed patient personality using the Big Five Inventory, a 44-item measure that generates scores on the five fixed personality dimensions: agreeableness, conscientiousness, extraversion, neuroticism, and openness.¹⁵ Personality was previously associated with patient satisfaction with outpatient healthcare.¹⁶ To minimize respondent burden, we carried initial responses to these three items forward to subsequent surveys for 138 patients attending 178 visits (13.5% of all visits).

Analyses. Analyses were conducted using Stata Version 14.2 (College Station, TX). Because patient experience scores were

highly skewed, we transformed the scores into percentile rank of visit (ranging from the worst visit rank of 0 to the best rank of 100).¹⁷ We fit multi-level linear mixed regression models with visit experience percentile rank as the dependent variable to permit the estimation of variance components for use in estimating intraclass correlation coefficients.^{18,19} Patient- and physician-level random effects modeled patients nested within physician and physicians cross-nested within patients. The base model included no covariates; the second model included variables recommended by the CAHPS Consortium for standard case-mix adjustment (i.e., age, gender, race/ethnicity, educational level, self-rated physical and mental health); finally, the expanded model added additional patient characteristics (medical skepticism, personality, global life satisfaction, marital status, BMI, symptom bother, and having seen the physician previously). For each model, we derived the patient- and physician-level intraclass correlation coefficients (ICCs) by dividing the between-patient and between-physician variance components, respectively, by the sum of all variance components (between-physician, between-patient, and residual error).^{18,20} In essence, ICCs quantify the size of the patient and physician effects on patient experience; a physician-level ICC that is much lower than the patient-level ICC implies that stable physician-level factors (e.g., communication skills) explain a much lesser amount of variation in patient experience than patient-level factors (e.g., socioeconomic, attitudes, personality). We obtained similar physician-level ICCs when we repeated the models using the raw patient experience z-scores so we report only results based on analyses of ranked z-scores. Because total variance explained is not provided within a cross-nested, multi-level model, we also performed linear regression using generalized estimating equations to estimate the total variation in patient experience (expressed as R²) explained by the two models with covariate adjustment (CAHPS Consortium standard adjustment and expanded adjustment).

We assessed the impact of greater adjustment on the required sample sizes to reliably detect the physician effect on patient experience. A lower physician-level ICC implies that a larger sample of patient surveys will be required to reliably measure the physician effect, because a greater proportion of the variation in the measures is explained by non-physician factors or is unexplained. Based on ICCs estimated with standard and expanded adjustment for patient-level factors, we used the Spearman-Brown prophecy formula to estimate the sample sizes required to achieve physician-level patient experience measures with reliabilities of 0.70, 0.80, and 0.90.²¹

Lastly, we assessed the impact of additional adjustment on assessment of patient experience measures at the individual physician level. Using the mixed models fitted with standard CAHPS adjusters and the expanded covariate set, we derived best linear unbiased predictors (BLUPs) for study physician performance in terms of patient experience.²² BLUPs are also called shrinkage (or Bayesian) estimates because they adjust for the uncertainty of the estimates, shrinking estimates toward

the mean according to the level of uncertainty (primarily the number of patient response at the physician level). BLUPs for both models were used to assign physicians to performance quartiles (highest, upper middle, lower middle, and bottom). From the two performance quartile measures, we derived a reclassification index, assessing the proportion of physicians who moved into or out of the top and bottom quartiles with expanded as compared to standard adjustment.

RESULTS

Patient experience surveys were completed by 1141 patients after 1319 primary care visits with 56 PCPs (mean visits per patient: 1.2, range: 1–6; mean number of physicians within

Table 1 Patient Characteristics (N = 1141)*

Characteristic	
Age, years, mean (SD)	45.6 (16.2)
Sex, n (%)	
Male	374 (32.8)
Female	767 (67.2)
Race/ethnicity, n (%)	
White	553 (48.5)
Hispanic	252 (22.1)
Black	125 (11.0)
Asian	75 (6.6)
Other/multiple	93 (8.2)
Decline to state	43 (3.8)
Education, n (%)	
< High school/GED	39 (3.4)
High school/GED	163 (14.3)
Some college	418 (36.6)
College graduate	255 (22.3)
Some post-college	266 (23.3)
Physical health status, n (%)	
Poor	47 (4.1)
Fair	182 (16.0)
Good	429 (37.6)
Very good	350 (30.7)
Excellent	133 (11.7)
Mental Health Index-5, mean (SD) (range: 4–100)	73.1 (18.5)
Any prior visit with physician, n (%)	
No	661 (57.9)
Yes	480 (42.1)
Medical skepticism, mean (SD), (range: 1–5)	3.0 (0.7)
Big-Five Personality Dimensions, mean (SD) (range, each dimension: 1–5)	
Extraversion	3.4 (0.8)
Agreeableness	4.2 (0.6)
Conscientiousness	3.9 (0.6)
Neuroticism	2.7 (0.8)
Openness	3.8 (0.6)
Global Life Satisfaction, mean (SD), (range: 5–35)	25.3 (6.5)
Symptom bothersomeness/worry, mean (SD) (range: 3–15)	8.0 (2.8)
BMI, kg/m ² , mean (SD)	29.8 (7.4)
Marital status, n (%)	
Divorced	146 (12.8)
Married/domestic partnership	510 (44.7)
Member of an unmarried couple	113 (9.9)
Never married	275 (24.1)
Separated	32 (2.8)
Widowed	65 (5.7)

*Patient experience surveys were completed by 1141 patient attending 1319 visits (mean, 1.2 visits/patient). Table includes patient characteristics reported on first visit

Table 2 Change in Percentile Visit Rank in the Patient Experience Measure Associated with a Unit-Change in Patient-Level Covariates

Standard adjustment*		
Variable	PE (95% CI)	p-value
Age, years	0.4 (0.3, 0.6)	<.001
Sex		
Male	1.0 (ref)	
Female	0.1 (-2.6, 2.8)	0.97
Race/ethnicity		
White	1.0 (ref)	
Hispanic	2.4 (-1.7, 6.5)	0.24
Black	3.4 (-2.2, 8.9)	0.23
Asian	4.6 (-2.8, 11.9)	0.22
Other/multiple	4.2 (-1.1, 9.4)	0.12
Decline to state	0.6 (-8.7, 9.9)	0.90
Education		
< High school/GED	1.0 (ref)	
High school/GED	1.8 (-6.9, 10.6)	0.68
Some college	-1.3 (-9.8, 7.3)	0.77
College graduate	-8.1 (-17.4, 1.2)	0.09
Some post-college	-6.4 (-15.6, 2.8)	0.17
Physical health status		
Poor	1.0 (ref)	
Fair	-1.7 (-12.1, 8.7)	0.74
Good	-2.4 (-11.8, 7.0)	0.61
Very good	1.4 (-8.0, 10.9)	0.76
Excellent	6.4 (-4.5, 17.3)	0.24
Mental Health Index-5 (range: 4–100)	0.1 (0.0, 0.2)	0.15
Variance explained†		
R ²	9.6%	
Expanded adjustment*		
Variable	PE (95% CI)	p-value
Age, years	0.3 (0.2, 0.5)	<.001
Sex		
Male	1.0 (ref)	
Female	0.1 (-2.5, 2.7)	0.94
Race/ethnicity		
White	1.0 (ref)	
Hispanic	3.2 (-0.6, 6.9)	0.10
Black	3.5 (-1.0, 8.1)	0.13
Asian	3.5 (-2.3, 9.3)	0.24
Other/multiple	5.4 (0.2, 10.6)	0.04
Decline to state	2.1 (-5.7, 9.8)	0.60
Education		
< High school/GED	1.0 (ref)	
High school/GED	2.1 (-6.7, 10.8)	0.64
Some college	-0.5 (-8.7, 7.8)	0.91
College graduate	-5.5 (-14.4, 3.3)	0.22
Some post-college	-5.9 (-14.8, 3.0)	0.19
Physical health status		
Poor	1.0 (ref)	
Fair	-2.8 (-12.1, 6.6)	0.56
Good	-4.1 (-13.0, 4.8)	0.36
Very good	0.1 (-8.8, 8.9)	0.99
Excellent	4.6 (-5.0, 14.2)	0.34
Mental Health Index- 5 (range: 4–100)	0.0 (-0.1, 0.1)	0.88
Any prior visit with physician		
No	1.0 (ref)	
Yes	14.6 (11.4, 17.8)	<.001
Medical skepticism (range: 1–5)	-4.8 (-7.0, -2.5)	<.001
Big-Five Personality Dimensions (range, each dimension: 1–5)		
Extraversion	-0.4 (-2.3, 1.4)	0.65
Agreeableness	2.7 (-0.2, 5.5)	0.06
Conscientiousness	1.9 (-0.5, 4.3)	0.12
Neuroticism	1.9 (-0.5, 4.3)	0.11
Openness	0.7 (-2.1, 3.5)	0.62
Global Life Satisfaction (range: 5–35)	0.5 (0.2, 0.7)	<.001
Symptom bothersomeness (range: 3–15)	-1.1 (-1.8, -0.4)	0.001
BMI, kg/m ²	0.3 (0.1, 0.5)	0.001
Marital status		
Divorced (ref)	1.0 (ref)	
Married/domestic partnership	-1.5 (-6.2, 3.2)	0.53
Member of an unmarried couple	-1.3 (-7.7, 5.0)	0.68

(continued on next page)

Table 2. (continued)

Never married	-0.9 (-6.4, 4.6)	0.74
Separated	-10.5 (-19.8, -1.3)	0.03
Widowed	-9.6 (-15.8, -3.5)	0.003
Variance explained†		
R ²	20.0%	

Abbreviations: PE parameter estimate

*Standard adjustment includes covariates recommended for case-mix adjustment by the Consumer Assessment of Healthcare Providers and Systems (CAHPS) program, while expanded adjustment includes additional patient-level covariates

†R² was estimated using generalized estimating equations, while parameter estimates derive from mixed-effects linear regression that accounts for cross-nesting of patients within physicians

patients: 1.1, range 1–5). The average number of visits with each PCP was 23.6 (median 32, range: 1–56). In 746 of 1319 visits (56.6%), patients reported that the visit was their first with the PCP. Of the 56 PCPs, 45 were resident physicians (80%); the rest were attending physicians. After encounters, patients rated 71.4% of visits as either 9 or 10 (“top box”) on the doctor rating scale compared to 82% of patients visiting US family medicine physicians whose practices submitted data to the 2015 Clinician and Group CAHPS Database.²³ The visit-level patient experience z-score was higher on average and had lesser variability among attending physicians [mean = 0.16, standard deviation (SD) = 0.50] as compared to residents (mean = -0.05, SD = 0.73). Table 1 shows characteristics of the 1141 patients.

With standard adjustment for a limited range of patient covariates, older patient age was statistically significantly associated with a higher percentile patient experience rank [parameter estimate = 0.4 (95% CI: 0.3, 0.6)] (Table 2). Overall, the standard patient-level covariates explained approximately 10% of the variation in patient experience measures across visits (R²=9.6%). With adjustment for the expanded covariate set, several additional patient factors were statistically significantly associated with visit experience, including any prior visit with the study physician, medical skepticism, global life satisfaction, symptom bothersomeness, BMI, and marital status. The model with expanded adjustment for patient covariates explained over twice as much variance as the model with standard adjustment [R² = 20.0% vs. 9.6%, respectively; *F* (10, 55) = 18.28, *p* < 0.001 for expanded vs. standard adjustment].

Without any adjustment, the physician-level ICC was 6.4%, decreasing to 6.1% with standard adjustment for patient covariates (Table 3). With expanded adjustment for patient characteristics, the physician-level ICC diminished to 3.4%, over 12-fold smaller than the patient-level ICC of 45.0% in the unadjusted model. Based on the physician ICC with standard adjustment, sample sizes of 36, 62, and 138 patient surveys would be required to achieve reliabilities of 0.70, 0.80, and 0.90, respectively (Table 3). These required sample sizes were close to required sample sizes without adjustment for patient characteristics (34, 59, and 132, respectively). With the

Table 3 Sample Size Requirements to Achieve Desired Reliabilities of Physician-Level Performance on Patient Experience Measures Under Varying Levels of Adjustment for Patient Characteristics

Extent of adjustment for patient characteristics	Patient-level ICC (95% CI), %	Physician level ICC (95% CI), %	Sample size to achieve $\alpha = 0.70$ (95% CI)	Sample size to achieve $\alpha = 0.80$ (95% CI)	Sample size to achieve $\alpha = 0.90$ (95% CI)
No adjustment	45.0 (35.0, 55.0)	6.4 (2.6, 10.2)	34 (21, 88)	59 (35, 150)	132 (79, 338)
Standard adjustment*	41.6 (31.2, 52.0)	6.1 (2.4, 9.8)	36 (22, 94)	62 (37, 161)	138 (83, 362)
Expanded adjustment*	33.8 (21.1, 46.5)	3.4 (0.7, 6.2)	66 (36, 351)	113 (61, 602)	255 (137, 1355)

ICC intraclass correlation coefficient

*Standard adjustment includes covariates recommended for case-mix adjustment by the Consumer Assessment of Healthcare Providers and Systems (CAHPS) program, while expanded adjustment includes additional patient-level covariates

diminution in physician ICC that accompanied expanded adjustment for patient characteristics, required samples sizes for reliabilities of 0.70, 0.80, and 0.90 nearly doubled to 66, 113, and 255, respectively.

The effect of expanded adjustment on physician reclassification is shown in Table 4. Among the 56 study physicians, 8 (14.3%; 95% CI: 7.0%, 25.2%) were reclassified from the top or bottom quartile to the middle (or vice versa).

DISCUSSION

Within a large academic family practice, patient-level factors explained a much larger percentage of variance in patient experience measures than physician-level factors, including patient attitudinal and dispositional factors that are not routinely measured on widely used patient experience surveys. Because the measurable physician effect on patient experience diminishes with greater adjustment for patient factors, large sample sizes of patient surveys would be needed to reliably measure patient experience at the individual physician level. Such large sample sizes may be infeasible for practices to achieve. Adjustment for an expanded set of patient covariates also had a substantive impact on the ranking of physicians based on patient experience. This change in ranking implies that patient attitudinal and dispositional variables would confound a ranking of physicians using only measurement procedures recommended by the CAHPS Consortium.

In our sample, several patient characteristics explained substantial added variance in patient experience compared to

covariates used in standard adjustment for CAHPS surveys, including medical skepticism, global life satisfaction, BMI, symptom bothersomeness, and marital status. Each of these variables may be plausibly associated with patient attitudes or beliefs that may influence patients' perceptions of life experiences, including healthcare experiences. Altogether, the expanded covariate set nearly doubled the amount of explained variation in visit-level patient experience. Ideally patient experience surveys would be modified to assess these additional patient characteristics. However, longer patient experience surveys would augment respondent burden and potentially depress already low response rates to mailed surveys.²⁴⁻²⁶

A prior study observed a strong association between patient visit experience and whether patients had established relationships with PCPs.¹⁰ Current CAHPS surveys collect information on prior visits with physicians, although the CAHPS Consortium recommends neither sampling based on established vs. non-established status nor adjustment for prior visits with physicians.^{7,27} Given the observed strength of this association, case-mix adjustment of patient experience measures should routinely adjust for prior visits with the physician.

The CAHPS Consortium recommends that healthcare organizations obtain sample sizes of 50 surveys per physician to allow valid comparisons of patient experience measures between physicians. Support for this recommendation derives from studies indicating that a sample of 50 is sufficient to achieve reliabilities of 0.70 in patient experience measures that are either unadjusted²⁸⁻³⁰ or adjusted for a limited range of covariates.²⁵ Our results are consistent with these studies in showing that a sample size of 50 is likely to be sufficient to achieve a reliability of 0.70 with limited adjustment for patient

Table 4 Reclassification of Physician-Level Ranking Based on Average Patient Experience with Standard and Expanded Adjustment for Patient-Level Factors

Quartile of physician ranking	Expanded adjustment*			Total	
	Top quartile	Middle quartiles	Bottom quartile		
Standard adjustment*	Top quartile	12	2	0	14
	Middle quartile	2	24	2	28
	Bottom quartile	0	2	12	14
	Total	14	28	14	56

*Standard adjustment includes covariates recommended for case-mix adjustment by the Consumer Assessment of Healthcare Providers and Systems (CAHPS) program, while expanded adjustment includes additional patient-level covariates. Overall, 8 of 56 physicians were reclassified with expanded adjustment (14.3%, 95% CI: 7.0-25.2%)

covariates (Table 3). However, for valid comparisons of measures at the individual physician level, some have suggested that measures should have a reliability of at least 0.80,³¹ while others have argued for higher reliabilities if physicians are to be held to account for experience measures.³⁰ As we and others have shown,^{32–35} physician-level ICCs for patient experience measures are small compared to patient-level ICCs and diminish further with broader adjustment for patient factors. With full adjustment for the range of patient covariates, we found that a sample size of 255 patient surveys would be required to measure physician-level patient experience with a reliability of 0.90, approximately 66% greater than analogous estimates with minimal or no adjustment for patient characteristics.^{25,30} Such a large sample size per physician is probably infeasible for most health systems to collect, especially with a longer survey that includes measures of influential patient attitudes and disposition.

Without expanded adjustment for patient characteristics or with small sample sizes, error in physician classification may be inevitable. We demonstrated the potential for error in a reclassification analysis in which one in seven physicians were reclassified into or out of the top or bottom performance quartiles after full adjustment for a range of patient characteristics vs. standard adjustment with a limited range of adjusters recommended by the CAHPS program. Because we employed conservative Bayesian statistical procedures to shrink outlying or small-sample estimates toward the mean, this reclassification analysis primarily illustrates the potential for error due to physician-level confounding by patient characteristics that are currently not accounted for in the CAHPS surveys. Confounding by patient-level factors has also been shown to substantially impact comparisons of physician-level performance on Healthcare Effectiveness Data and Information Set (HEDIS) quality metrics.³⁶

Our visit-level patient experience outcome was a composite measure comprising six items, four of which derive from the CAHPS Physician Communication Composite. Compared to single-item measures (e.g., doctor rating), a scale derived from multiple items has greater measurement reliability.³⁰ For this reason, the CAHPS Consortium encourages the use of composite patient experience measures for public reporting of group- or physician-level measures.³⁷

Our study sample derives from a single academic family medicine practice. While results could differ in other settings, our minimally adjusted findings were consistent with prior reports.^{25,28–30} In addition, patients were seeing physicians for the first time in most visits (56.6%), reducing potential halo effects from prior experiences with physicians. The lack of continuity implied by the high proportion of first visits implies that patients with more than one visit typically saw more than one physician (essentially randomly determined by the scheduling process). The inclusion of both resident and faculty physicians introduced greater variation in physician interpersonal skills and consequent variation in patient experience. While these features may limit the generalizability of the findings to community practices, they enabled greater exploration of the extent to which

variance can be attributed to physician vs. patient factors. Because greater variability enhances the measure reliability, reliability would be expected to be lower in a practice with attending physicians only, and required sample sizes to allow meaningful comparison of individual physicians would likely be higher than we have estimated within this academic practice.

We recruited patients prior to visits to participate voluntarily in post-visit surveys and were unable to ascertain the number of patients who were aware of the opportunity or the number who opted not to participate. Although we suspect the overall response rate was low, response rates are also low to mailed patient experience surveys typically used to assess physician performance.^{24–26,29} Response bias may affect study results and the results of other patient experience surveys.³⁸ Because patients responded to questions on a tablet device rather than a mail survey, mode effects may also influence results.³⁹

Although patient experience surveys are commonly used to manage physicians and to determine their compensation,^{2,40} our results raise several questions about the validity and practicality of comparing individual primary care physicians based on patient experience scores. First, we found that the limited range of covariates typically adjusted for in physician-level comparisons did not account sufficiently for the variation between patients in attitudes, beliefs, and other patient characteristics that influence patients' perceptions of their healthcare experience. Second, while adjustment for a broader range of patient covariates augmented the explained variation in patient experience, it diminished the size of the physician effect on these measures, and sample size requirements to detect this effect increased proportionally. Most practices will lack the resources to collect such large samples of an expanded patient experience survey. Third, we demonstrated the potential for substantial confounding at the physician-level by unmeasured patient characteristics if practices compared physicians based on patient experience scores adjusted only for the limited range of patient measures recommended by the CAHPS Consortium. Although our data were derived from a single practice, these results suggest that healthcare organizations should exercise caution when comparing individual primary care physicians on average measures of patient experience using current measures.

Acknowledgements: The authors thank Rima Cabrera, MSW, for coordinating the project and Eliot Lee and Leyleh Salem for assistance in data collection. Dr. Fenton had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs. Franks, Tancredi, and Fenton conducted and were responsible for the data analyses.

Corresponding Author: Joshua J. Fenton, MD, MPH; Department of Family and Community Medicine University of California, Davis Health System, Sacramento, CA, USA (e-mail: jifenton@ucdavis.edu).

Compliance with Ethical Standards:

Conflicts of Interest: The authors declare no conflicts of interest.

Financial Support: This work was supported by the Department of Family and Community Medicine, University of California, Davis. Dr. Magnan was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through grant #UL1

TR001860 and linked award KL2 TR001859. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Several authors are faculty members in the Department of Family and Community Medicine (Fenton, Jerant, Bertakis, Magnan, and Franks). All authors were involved in the design and conception of the project, the collection and interpretation of the study data, the composition and editing of the research manuscript, and the decision to submit it for publication.

REFERENCES

- Anhang Price R, Elliott MN, Cleary PD, Zaslavsky AM, Hays RD. Should health care providers be accountable for patients' care experiences? *J Gen Intern Med.* 2015;30(2):253-256.
- Ding VY, Hubbard RA, Rutter CM, Simon GE. Assessing the accuracy of profiling methods for identifying top providers: performance of mental health care providers. *Health Serv Outcome Res Methodol.* 2013;13(1):1-17.
- Browne K, Roseman D, Shaller D, Edgman-Levitan S. Analysis & commentary. Measuring patient experience as a strategy for improving primary care. *Health Aff (Millwood).* 2010;29(5):921-925.
- Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA.* 1999;281(22):2098-2105.
- Nyweide DJ, Weeks WB, Gottlieb DJ, Casalino LP, Fisher ES. Relationship of primary care physicians' patient caseload with measurement of quality and cost performance. *JAMA.* 2009;302(22):2444-2450.
- Agency for Healthcare Research and Policy. CAHPS: Surveys and tools to advance patient care. 2016; <http://www.ahrq.gov/cahps/surveys-guidance/cg/index.html>. Accessed 11 November, 2016.
- CAHPS Surveys and Instructions. Instructions for analyzing data from CAHPS surveys: Using the CAHPS analysis program version 4.1. Agency for Healthcare Quality and Research; Document No. 2015v2. 2012.
- Berwick DM, Murphy JM, Goldman PA, Ware JE, Jr., Barsky AJ, Weinstein MC. Performance of a five-item mental health screening test. *Med Care.* 1991;29(2):169-176.
- Hewitt B, Turrell G, Giskes K. Marital loss, mental health and the role of perceived social support: findings from six waves of an Australian population based panel study. *J Epidemiol Community Health.* 2012;66(4):308-314.
- Rodriguez HP, Rogers WH, Marshall RE, Safran DG. The effects of primary care physician visit continuity on patients' experiences with care. *J Gen Intern Med.* 2007;22(6):787-793.
- Hinchey SA, Jackson JL. A cohort study assessing difficult patient encounters in a walk-in primary care clinic, predictors and outcomes. *J Gen Intern Med.* 2011;26(6):588-594.
- Fong RL, Bertakis KD, Franks P. Association between obesity and patient satisfaction. *Obesity.* 2006;14(8):1402-1411.
- Fiscella K, Franks P, Clancy CM. Skepticism toward medical care and health care utilization. *Med Care.* 1998;36(2):180-189.
- Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. *J Pers Assess.* 1985;49(1):71-75.
- Digman JM. Personality structure: emergence of the five-factor model. *Annu Rev Psychol.* 1990;41:417-440.
- Costello BA, McLeod TG, Locke GR, 3rd, Dierkhising RA, Offord KP, Colligan RC. Pessimism and hostility scores as predictors of patient satisfaction ratings by medical out-patients. *Int J Health Care Qual Assur.* 2008;21(1):39-49.
- Conover WJ, Iman RL. Rank transformations as a bridge between parametric and nonparametric statistics. *Am Stat.* 1981;35(3):124-129.
- Carrasco JL, Jover L. Estimating the generalized concordance correlation coefficient through variance components. *Biometrics.* 2003;59(4):849-858.
- Kraemer HC. Correlation coefficients in medical research: from product moment correlation to the odds ratio. *Stat Methods Med Res.* 2006;15(6):525-545.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86(2):420-428.
- Carmines EG, Zeller RA, eds. Reliability and validity assessment. In: MS L-B, ed. Sage University Paper Series on Quantitative Applications in the Social Sciences. Newbury Park, Calif.: Sage Publications; 1979. No. Series no. 07-001.
- Datta GS, Ghosh M. Bayesian prediction in linear models: Applications to small area estimation. *Ann Stat.* 1991;19(4):1748-1770.
- Shaller D, Yount N, Rauch J, Li S, Corrothers M. 2015 Chartbook: What Patients Say About Their Healthcare Providers and Practices. Rockville, MD: Agency for Healthcare Policy and Research; 2016.
- Rodriguez HP, von Glahn T, Chang H, Rogers WH, Safran DG. Measuring patients' experiences with individual specialist physicians and their practices. *Am J Med Qual.* 2009;24(1):35-44.
- Safran DG, Karp M, Coltin K, et al. Measuring patients' experiences with individual primary care physicians. Results of a statewide demonstration project. *J Gen Intern Med.* 2006;21(1):13-21.
- Selby JV, Schmittiel JA, Lee J, et al. Meaningful variation in performance: what does variation in quality tell us about improving quality? *Med Care.* 2010;48(2):133-139.
- Fielding the CAHPS Clinician & Group Survey. In CAHPS Clinician & Group Survey Instructions. Bethesda, MD: Agency for Healthcare Quality and Research; Document No. 2033, 2016.
- Dyer N, Sorra JS, Smith SA, Cleary PD, Hays RD. Psychometric properties of the Consumer Assessment of Healthcare Providers and Systems (CAHPS(R)) Clinician and Group Adult Visit Survey. *Med Care.* 2012;50 Suppl:S28-34.
- Hays RD, Chong K, Brown J, Spritzer KL, Horne K. Patient reports and ratings of individual physicians: an evaluation of the DoctorGuide and Consumer Assessment of Health Plans Study provider-level surveys. *Am J Med Qual.* 2003;18(5):190-196.
- Nelson EC, Gentry MA, Mook KH, Spritzer KL, Higgins JH, Hays RD. How many patients are needed to provide reliable evaluations of individual clinicians? *Med Care.* 2004;42(3):259-266.
- McDowell I. *Measuring Health: A Guide to Rating Scales and Questionnaires.* 3rd ed. New York, NY: Oxford University Press; 2006.
- Franciosi M, Pellegrini F, De Berardis G, et al. Correlates of satisfaction for the relationship with their physician in type 2 diabetic patients. *Diabetes Res Clin Pract.* 2004;66(3):277-286.
- McKinley RK, Roberts C. Patient satisfaction with out of hours primary medical care. *Qual Health Care.* 2001;10(1):23-28.
- Sixma HJ, Spreeuwenberg PM, van der Pasch MA. Patient satisfaction with the general practitioner: a two-level analysis. *Med Care.* 1998;36(2):212-229.
- Bjorngaard JH, Ruud T, Garratt A, Hatling T. Patients' experiences and clinicians' ratings of the quality of outpatient teams in psychiatric care units in Norway. *Psychiatr Serv (Washington, D.C.).* 2007;58(8):1102-1107.
- Hong CS, Atlas SJ, Chang Y, et al. Relationship between patient panel characteristics and primary care physician clinical performance rankings. *JAMA.* 2010;304(10):1107-1113.
- Consumer Assessment of Healthcare Providers and Systems. Patient experience measures from the CAHPS clinician and group surveys. Agency for Healthcare Quality and Research; Document No. 1309, 2014.
- Groves RM. Nonresponse rates and nonresponse bias in household surveys. *Pub Opin Q.* 2006;70(5):646-675.
- Elliott MN, Zaslavsky AM, Goldstein E, et al. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res.* 2009;44(2 Pt 1):501-518.
- Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *JAMA.* 2007;297(7):740-744.