

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<https://hdl.handle.net/2066/227437>

Please be advised that this information was generated on 2021-06-15 and may be subject to change.

Measuring Patient-Reported Shared Decision-Making to Promote Performance Transparency and Value-Based Payment: Assessment of collaboRATE's Group-Level Reliability

Journal of Patient Experience
2020, Vol. 7(5) 742-748
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2374373519884835
journals.sagepub.com/home/jpx


Rachel C Forcino, MSc¹ , Marcus Thygeson, MD MPH²,
A James O'Malley, PhD^{1,3}, Marjan J Meinders, PhD⁴,
Gert P Westert, PhD⁴, and Glyn Elwyn, MD PHD¹ 

Abstract

Shared decision-making (SDM) between clinicians and patients is a key component of patient experience, but measurement efforts have been hampered by a lack of valid and reliable measures that are feasible for routine use. In this study, we aim to investigate **collaboRATE**'s reliability, calculate required sample sizes for reliable measurement, and compare Consumer Assessment of Healthcare Providers and Systems (CAHPS) patient experience survey items to **collaboRATE**. **CollaboRATE**'s provider group-level reliability reached acceptable reliability at 190 patient reports, while the CAHPS SDM measure demonstrated similar reliability at a sample size of 124. The CAHPS communication measure reached acceptable reliability with 55 patient reports. A strong correlation was observed between **collaboRATE** and CAHPS communication measures ($r = 0.83$). As a reliable measure of SDM, **collaboRATE** may be useful for both building payment models that support shared clinical decision-making and encouraging data transparency with regard to provider group performance.

Keywords

communication, measurement, medical decision-making, outpatient satisfaction data, patient feedback, patient satisfaction

Background

Given increasing reliance on patient experience measures in US performance-based compensation systems, as well as evolving public expectations with regard to patient-centered health care, a focus on patient experience is critical (1). Patient experience measurement has been described as “reports from patients on what they did or did not experience in their interactions with providers and the healthcare system” (2). In addition to the importance of positive patient experience per se, there is increasing evidence of positive correlations between experience, patient safety, and clinical outcomes (3). Improving the measurement and transparency of patients' experience could contribute widely to improvement in the quality of healthcare delivery.

As a key component of patient experience, shared decision-making (SDM) between clinicians and patients has garnered increased attention in recent years, having been promoted in the US Patient Protection and Affordable

Care Act (4,5) and championed as “the pinnacle of patient-centered care” (6). However, the ability to assess patients' experience of SDM has been hampered by a lack of measures that have proven psychometric properties such as validity, responsiveness, and, critical for provider group performance assessment, reliability, while at the

¹ The Dartmouth Institute for Health Policy & Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

² Bind Benefits, Minneapolis, MN, USA

³ Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

⁴ Scientific Center for Quality of Healthcare, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, NL, USA

Corresponding Author:

Rachel C Forcino, The Dartmouth Institute, Williamson Level 5, 1 Medical Center Drive, Lebanon, NH 03756, USA.
Email: rachel.forcino@dartmouth.edu



same time are practical enough for use in routine clinical settings. This measurement gap also hinders SDM performance improvement.

To help meet this need, Elwyn and colleagues developed **collaboRATE**, a 3-item patient-reported measure of SDM process that applies to any healthcare decision, has strong face validity among its target population, and is possible to complete in less than 30 seconds (7). Preliminary psychometric testing of **collaboRATE** in a simulated, online sample has established its concurrent validity with existing research-oriented patient-reported measures of SDM such as the SDM-Q-9 questionnaire (8,9). A feasibility study in 3 US primary care practices has demonstrated its practicable use in routine ambulatory healthcare service (10).

Another approach to SDM measurement was undertaken by the US Agency for Healthcare Research and Quality as a supplement to the widespread Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group Survey (CG-CAHPS version 3.0). The CAHPS survey is used by providers and payers of US health-care services to “evaluate and compare healthcare providers and to improve the quality of healthcare services” (11). The CAHPS supplemental SDM section focuses on medication-related decisions—discussing reasons to take medication, discussing reasons not to take medication, and discussing which medication is best. The CAHPS communication items evaluate clinician performance in explaining things in a way that is easy to understand, listening carefully, showing respect, and spending enough time with the patient.

Although CAHPS scores have been evaluated in various healthcare settings, no prior studies have compared CAHPS communication and SDM measures to **collaboRATE** on key elements of provider group performance measurement. In the current study, we therefore aim to (1) report on provider group performance on **collaboRATE** administered in a CAHPS-like survey, (2) compare required sample sizes for reliable provider group performance profiling using **collaboRATE** to those required for CAHPS patient experience survey items relating to clinician communication and SDM, and (3) evaluate **collaboRATE**'s concurrent validity with CAHPS clinician communication and SDM composite scores.

Methods

Data Sources

We conducted a secondary analysis of the full cross-sectional 2017 California Patient Assessment Survey data set. The Patient Assessment Survey is administered annually by the Pacific Business Group on Health to adult patients of 153 California provider groups taking commercial capitation risk, and results are incorporated into California's pay-for-performance program. Participating provider groups are geographically diverse practices representing all of California's major metropolitan regions as well as rural areas within the

state. Each provider group includes both primary and specialty care clinicians.

Our use of the deidentified 2017 Patient Assessment Survey data set was approved by the Pacific Business Group on Health. This project was considered exempt from further review by Dartmouth College's Committee for the Protection of Human Subjects (study #31002).

Participants. Patient Assessment Survey participants include privately insured adult (aged 18 years and older) patients who received ambulatory healthcare services at one of the 153 participating California provider groups between January and October 2016. The Patient Assessment Survey's standard sampling procedure excludes patients younger than 18 years of age, uninsured patients, and patients with public or other noncommercial health insurance. Therefore, our secondary analysis of the Patient Assessment Survey data set also excludes these populations.

Questionnaire administration. In accordance with the standard Patient Assessment Survey administration procedure, a random sample of eligible patients were invited to complete the survey. The recruitment protocol included a series of survey administration modes, beginning with e-mail invitations and progressing to mail then attempting telephone administration for prior nonrespondents. All questionnaires were administered between December 2016 and March 2017. Questionnaires were available in English, Spanish, Chinese, Korean, and Vietnamese languages.

Measures. The 2017 Patient Assessment Survey consisted of 35 total items derived from (1) the CAHPS Clinician & Group version 3.0 survey and supplement (12) and (2) the **collaboRATE** measure (7). This study focused on 10 items related to clinician communication and SDM, listed in Table 1.

Consumer Assessment of Healthcare Providers and Systems SDM item responses were given on a yes/no response scale, while CAHPS communication items were rated on a never/sometimes/usually/always response scale. **CollaboRATE** items were rated on an ordinal scale from 0, labeled “Worst care possible,” to 10, labeled “Best care possible.”

Statistical Analysis

In all analyses, survey responses with missing data on the outcome measure of interest were excluded from the analysis. Due to our interest in evaluating reliability under a worst-case scenario that emphasizes the relation of missing data to reliability, we did not use multiple imputation. Data were unweighted and physician group scores were case-mix adjusted by patient age, education, general and mental health status, race/ethnicity, mode and language of survey administration, and provider specialty. Analysis was conducted using Stata 13 statistical software.

Table 1. Shared Decision-Making and Communication Survey Items.

collaboRATE Items	CAHPS SDM Items	CAHPS Communication Items
How much effort did this doctor make to help you understand your health issues?	Did you and this doctor talk about the reasons you might want to take medicine?	How often did this doctor explain things in a way that was easy to understand?
How much effort did this doctor make to listen to the things that matter most to you about your health issues?	Did you and this doctor talk about the reasons you might not want to take medicine?	How often did this doctor listen carefully to you?
How much effort did this doctor make to include what matters most to you in choosing what to do next?	When you and this doctor talked about starting or stopping a prescription medicine, did this doctor ask what you thought was best for you?	How often did this doctor show respect for what you had to say?
—	—	How often did this doctor spend enough time with you?

Abbreviations: CAHPS, Consumer Assessment of Healthcare Providers and Systems; SDM, shared decision-making.

Scoring methods. For each of the 3 measures (**collaboRATE**, CAHPS communication, and CAHPS SDM), item scores were calculated by provider group as the proportion of all responses in which a top score (ie, “10” for **collaboRATE** items, “Always” for CAHPS communication items, and “Yes” for CAHPS SDM items) was given. Overall top-box scores represent the proportion of responses in which top scores were given on all items in the measure. Composite scores were calculated as the mean of the measure’s item scores. We scored each measure according to its customary approach: previously validated top-box scoring was adopted for **collaboRATE** (9), while CAHPS mean scoring was used for CAHPS communication and SDM composites. Ceiling effects, where responses tend toward the upper extreme of a measurement scale, are common among patient-reported experience measures such as **collaboRATE**. To mitigate these ceiling effects, and in accordance with the standard scoring procedures for each included measure, we followed the precedent specified by Barr et al (9) and adopted top-box scoring in which the highest possible score of 10 constituted the top box.

Provider group-level reliability analysis. Score reliability is a statistical measure of “how well one can confidently distinguish the performance of one physician [or provider group] from another” (13). The reliability of provider group performance

Table 2. Snijders and Bosker (1999) Reliability Formula.

$$\text{reliability} = \frac{n \times \rho}{1 + (n-1)\rho}$$

scores depends on 2 things: “(1) a sufficient number of patients eligible for a given quality measure ([eg, SDM] per provider group) and (2) performance variation across [provider groups] on that quality measure (14).” To evaluate provider group-level score reliability, we adopted the approach used by Scholle and detailed by Adams (13,14). For each measure (CAHPS communication, CAHPS SDM, and **collaboRATE**), we estimated a mixed effects logistic regression model and the intraclass correlation coefficient (ICC) implied by the model. We then calculated provider group-level score reliability using the procedure specified by Adams (13) and the formula detailed by Snijders and Bosker (15), reproduced in Table 2, where ρ is the case-mix adjusted ICC and n is the provider group-level sample size. This expression is derived from the definitions $\rho = \frac{\tau^2}{\tau^2 + \sigma^2}$ and reliability $\frac{\tau^2}{\tau^2 + \sigma^2/n}$, where τ^2 denotes the between provider group variance and σ^2 is the within provider group (between patient) variance, showing that reliability equals ICC when $n = 1$ and otherwise exceeds the ICC.

We report ICCs and median reliability estimates for each measure and across all provider groups (14). As reliability depends heavily on sample sizes, we also report response rates for each measure. Finally, we conducted a sensitivity analysis to determine the impact of the patient case-mix adjustment on provider group-level score reliability.

A minimum threshold for acceptable score reliability for performance profiling at the provider group level has been established at 0.70 (13,14). We therefore present the minimum number of patient reports required at the provider group level to reach the minimum 0.70 reliability threshold for each included measure.

Concurrent validity analysis. To examine concurrent validity between CAHPS communication, CAHPS SDM, and **collaboRATE** measures, we conducted Pearson correlation analysis comparing provider group-level scores by individual item as well as by composite and top-box score.

Limitations

Our study design had several limitations. First, as per Patient Assessment Survey standard procedures, questionnaires were administered to patients more than 2 months following their clinic visits, presenting the possibility of recall bias. Further, the 10 survey items analyzed in this study were among 28 total items contained within the 2017 Patient Assessment Survey; the impact of those additional 18 items on possible selection and response biases is unknown. Finally, we have access to demographic data only for respondents to the Patient Assessment Survey; we do not have access to detailed data on their respective healthcare

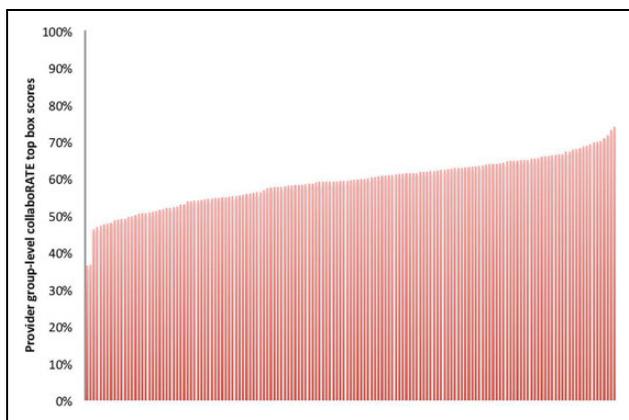


Figure 1. Case-mix adjusted collaboRATE top-box scores by provider group (n = 153 groups).

Table 3. CAHPS Communication and Shared Decision-Making Provider Group Scores.

Measure/Item	Score Range Across Provider Groups
Communication composite	65.2%-91.4%
CAHPS “explain”	66.3%-92.1%
CAHPS “listen”	64.9%-97.1%
CAHPS “respect”	70.3%-94.9%
CAHPS “time”	57.9%-88.8%
SDM composite	74.2%-94.7%
CAHPS “start medication”	88.0%-99.2%
CAHPS “stop medication”	60.5%-91.7%
CAHPS “best medication”	69.3%-93.7%

Abbreviations: CAHPS, Consumer Assessment of Healthcare Providers and Systems; SDM, shared decision-making.

providers or provider groups. We are therefore unable to further contextualize our results based on these individual provider and group characteristics.

Results

A demographic profile of patient respondents from 153 participating provider groups is available in Online Appendix 1.

CollaboRATE and CAHPS scores

The 2017 Patient Assessment Survey included 31 265 patient responses across 153 California provider groups. **CollaboRATE** scores varied by provider group, with top-box scores ranging from 36.7% to 74.8%. The overall mean **collaboRATE** score was 59.4%. Figure 1 displays the distribution of **collaboRATE** scores by provider group. The CAHPS communication composite scores ranged from 60.4% to 91.8% across provider groups, and CAHPS SDM composite scores ranged from 74.2% to 94.6%. Group-level scores for each CAHPS communication and SDM composite and individual item are listed in Table 3.

Provider Group-Level Score Reliability

Based on case-mix adjusted scores, **collaboRATE** provider group-level reliabilities had a median value of 0.71 (at $n = 204$; reliability range: 0.28-0.93 with n ranging 31-1133), while the median CAHPS SDM group-level reliability (at $n = 105$) was 0.67 (range: 0.23-0.92 with n ranging 16-614). Group-level reliability of CAHPS communication items had a median value of 0.90 (range: 0.57-0.98 with n ranging 31-1127). Reliability results are summarized in Table 4.

Given its ICC of 0.01, **collaboRATE** reaches 0.70 reliability with 190 patient reports. The CAHPS communication composite, with an ICC of 0.04, reaches 0.70 reliability with 55 patient reports and the CAHPS SDM composite, with an ICC of 0.02, reaches 0.70 reliability with 124 patient reports.

CollaboRATE had a 100% response rate among the 31 265 total patient respondents, while the CAHPS communication composite was answered by 99.6% of respondents ($n = 31 129$) and the CAHPS SDM composite was answered by 52.6% of respondents ($n = 16 460$).

In sensitivity analysis of scores unadjusted for patient case mix, the CAHPS communication and SDM group-level composite scores shifted an average of 11.5% and 16.1%, respectively, following case-mix adjustment; **collaboRATE** scores remained consistent with, on average, only a 0.2% change in group-level scores. CAHPS communication group-level scores had a range of 65.2% to 91.4% (mean = 82.7%) after case-mix adjustment, compared with a range of 44.1% to 85.4% (mean = 83.2%) preadjustment. Similarly, CAHPS SDM scores ranged 74.2% to 94.7% (mean = 86.4%) after case-mix adjustment and 50% to 87.7% (mean = 87.0%) preadjustment. **CollaboRATE** scores ranged 36.7% to 74.8% (mean = 59.4%) after case-mix adjustment and 36.5% to 75.0% (mean = 60.0%) preadjustment, suggesting little change in the inter-group variance after case-mix adjustment of **collaboRATE** scores.

Concurrent Validity

At the measure level, the strongest association was observed between **collaboRATE** and the CAHPS communication composite ($r = 0.83$). Correlations were moderate between the CAHPS SDM composite and **collaboRATE** ($r = 0.52$) and the CAHPS SDM and communication composites ($r = 0.61$). Figure 2 shows correlations between **collaboRATE** and CAHPS at both composite and item levels.

Between individual items, strong associations were observed among the **collaboRATE** items themselves; this correlation was especially pronounced between the **collaboRATE** “listen” and “help you understand” items ($r = 0.94$). The **collaboRATE** “listen” item was also highly correlated with the CAHPS “spend time” ($r = 0.83$), “listen” ($r = 0.82$), and “respect” ($r = 0.81$) items. The 3 CAHPS SDM items had low-to-moderate correlations with the 3 **collaboRATE** items (range: $r = 0.22$ -0.53).

Table 4. Provider Group-Level Reliability.

Measure	Median Sample Size (n)	ICC (ρ)	Median Reliability	Reliability Range (n = 153 Provider Groups)
Adjusted scores				
collaboRATE	204	0.01	0.71	0.28-0.93
CAHPS SDM	105	0.02	0.67	0.23-0.92
CAHPS communication	204	0.04	0.90	0.57-0.98
Unadjusted scores				
collaboRATE	204	0.01	0.74	0.30-0.94
CAHPS SDM	105	0.03	0.79	0.36-0.96
CAHPS communication	204	0.05	0.91	0.61-0.98

Abbreviations: CAHPS, Consumer Assessment of Healthcare Providers and Systems; ICC, intraclass correlation coefficient; SDM, shared decision-making.

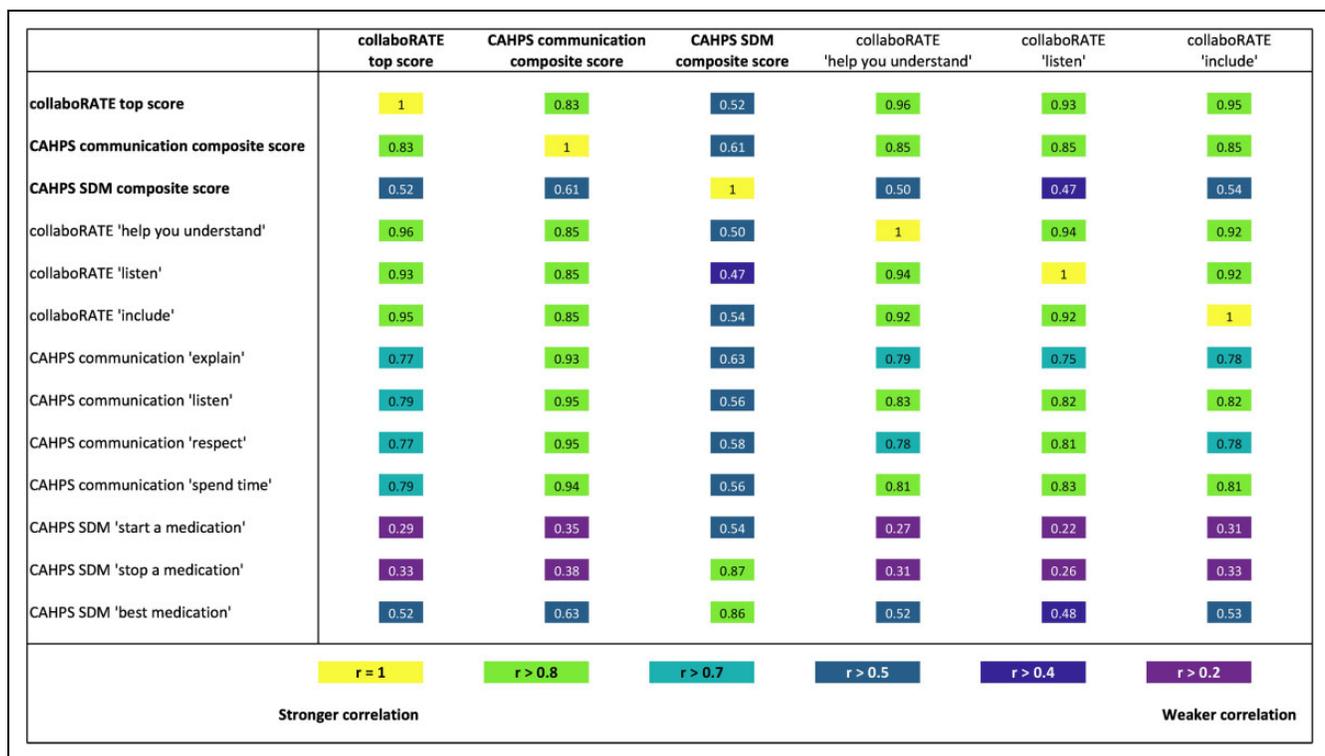


Figure 2. CollaboRATE correlations with Consumer Assessment of Healthcare Providers and Systems (CAHPS) communication and shared decision-making measures.

Discussion

Key Findings

Although **collaboRATE** shows a range of scores across the 153 provider groups represented in this sample, ICCs for all 3 included measures were low, meaning that relatively large sample sizes are needed to reach adequate levels of reliability for performance profiling applications. **CollaboRATE** demonstrates adequate reliability for performance profiling (13) with a minimum of 190 patient responses per provider group, the CAHPS communication composite reaches adequate reliability with a minimum of 55 patient responses per provider group, and the CAHPS SDM composite reaches adequate reliability with a minimum of 124 patient responses per provider group.

Correlation analysis demonstrates limited association between the **collaboRATE** measure and the CAHPS SDM composite, indicating a lack of concurrent validity between these 2 measures of SDM. Correlations are substantially higher between **collaboRATE** and CAHPS communication measures. Item-level analysis showed that CAHPS communication items are highly associated with **collaboRATE** items—an intuitive relationship between a measure of SDM and a measure of closely related communication practices including explaining, listening, showing respect, and spending enough time. It is possible that a provider's engagement in SDM may, in fact, result in higher communication ratings. Despite this high correlation between CAHPS communication and **collaboRATE** measures, there is insufficient evidence to conclude that the CAHPS

communication composite adequately captures SDM performance. Instead, a lack of face validity caused by exclusion of a key element of SDM, namely including patients' preferences in treatment planning or other next steps, calls into question the validity of the CAHPS communication composite as a proxy for SDM performance (7,16).

As the CAHPS SDM measure specific to medication decision-making is relevant to only the fraction of clinical encounters in which patients and providers make explicit decisions about starting or stopping a medication (no more than 52.6% of respondents in this sample), sample sizes for the CAHPS SDM measure are consistently low; these smaller sample sizes (median $n = 105$ across 153 provider groups compared to median $n = 204$ for **collaboRATE** and CAHPS communication measures) hinder the measure's reliability. In this case, the ability of the measure to detect true differences between provider groups is compromised—even when there is just as much true variation. This issue of limited relevance to a variety of clinical encounters highlights the need for a generic and broadly applicable measure of SDM such as **collaboRATE** to be collected on a routine basis. Further, as the third **collaboRATE** item (“include”) captures the definitive element of SDM (7,16,17), the low correlation between the **collaboRATE** “include” item and CAHPS SDM composite score suggests there is a construct mismatch inherent to the CAHPS SDM measure. Established concurrent validity between **collaboRATE** and SDM-Q-9, a validated research measure of SDM (9), and the lack of concurrent validity between **collaboRATE** and the CAHPS SDM composite demonstrated in this study together indicate that the CAHPS SDM composite lacks construct validity as a measure of SDM performance. This lack of construct validity, paired with inadequate reliability for performance measurement at the provider group level, suggests that the CAHPS SDM composite is an inadequate group-level measure of SDM performance.

Context in Existing Literature

While a strict reliability standard of 0.70 ensures fair comparison across provider groups for national or regional performance management and incentivization purposes, we recognize that obtaining the required sample sizes to meet this reliability standard may be too resource-intensive to be attainable in local quality improvement and research projects. For these local projects, we therefore recommend 20 to 30 observations per provider group in analyses designed to assess differences between groups and approximately 50 observations per provider group for inferential analyses (18).

Our findings support Hays' evaluation of CAHPS measure reliability, which found larger required sample sizes for the CAHPS SDM composite ($n = 396$) than for the CAHPS communication composite ($n = 295$) (19). However, that study's use of the Spearman-Brown reliability formula (20) resulted in larger recommended sample sizes than those we report in the current study (CAHPS communication:

$n = 55$; CAHPS SDM: $n = 124$) which used Adams' estimation approach (13).

We observed consistent mean **collaboRATE** scores both pre- and post-case-mix adjustment, with a substantial upward post-adjustment shift in group-level scores occurring primarily at the low end of the range. This is consistent with prior research finding that “case-mix adjustment of practice-level scores results in relatively few large adjustments (which were mainly positive), and many small adjustments (which were more often negative)” (21). Patient-level frequencies show that the majority of respondents self-identify as Asian in the 5 provider groups with the lowest unadjusted **collaboRATE** scores. More research is needed to explore the patient-level predictors of **collaboRATE** scores.

Adjusting for case mix is often advocated in healthcare quality measurement (21). It is likely that SDM scores will vary because of patient characteristics; language barriers, literacy levels, and other patient sociodemographic factors may limit perceived (or actual) SDM. Case-mix adjustment parses the variance attributable to provider performance from the variance associated with patient characteristics. Hence, as case-mix adjustment reduces overall variation between measured entities (22), the reliability measures based on case-mix adjusted scores may represent lower bound estimates, while the reliability measures calculated using unadjusted scores can be considered upper bound estimates.

Conclusion

Valid measurement relies on clarity about the constructs being measured, as well as meaningful interpretation of variation in scores. In this study, we demonstrate that **collaboRATE** has adequate reliability for provider group performance profiling with a minimum sample size of 190 patient reports. With this minimum sample size and given its fidelity to the core dimensions of SDM, **collaboRATE** could be considered as a group-level SDM performance measure. Smaller samples may be useful for quality improvement initiatives. Further research is needed to evaluate the relationship between patient case mix and patients' experience of SDM, and to investigate **collaboRATE**'s reliability as a measure of individual clinician performance.

Acknowledgments

The authors gratefully acknowledge Grace Lin, MD, for facilitating PAS data access.

Author Contributions

Rachel C Forcino, Marcus Thygeson, Glyn Elwyn contributed to conception or design of the work. Rachel C Forcino, Marcus Thygeson, A James O'Malley, and Glyn Elwyn contributed to the acquisition, analysis, or interpretation of data. Rachel C Forcino drafted the manuscript; Marcus Thygeson, A James O'Malley, Marjan J Meinders, Gert P Westert, and Glyn Elwyn critically revised the manuscript.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Glyn Elwyn is a developer of the **collaboRATE** measure; he is also an adviser to PatientWisdom, an organization which offers to collect **collaboRATE** data.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Rachel C Forcino  <https://orcid.org/0000-0001-9938-4830>

Glyn Elwyn  <https://orcid.org/0000-0002-0917-6286>

Supplemental Material

Supplemental material for this article is available online.

References

- Agency for Healthcare Research & Quality. Section 2: Why Improve Patient Experience?. Retrieved April 25, 2018, from: <https://www.ahrq.gov/cahps/quality-improvement/improvement-guide/2-why-improve/index.html>.
- Browne K, Roseman D, Shaller D, Edgman-Levitan S. Measuring patient experience as a strategy for improving primary care. *Health Aff (Millwood)*. doi:10.1377/hlthaff.2010.0238.
- 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:e001570.
- 111th United States Congress. The patient protection and affordable care act. Public Law. 2010;111-48.
- Oshima Lee E, Emanuel EJ. Shared decision making to improve care and reduce costs. *N Engl J Med*. 2013;368:6-8.
- Barry MJ, Edgman-Levitan S. Shared decision making—the pinnacle of patient-centered care. *N Engl J Med*. 2012;366:780-1.
- Elwyn G, Barr PJ, Grande SW, Thompson R, Walsh T, Ozanne EM. Developing collaborate: a fast and frugal patient-reported measure of shared decision making in clinical encounters. *Patient Educ Couns*. 2013;93:102-7.
- Kriston L, Scholl I, Hölzel L, Simon D, Loh A, Härter M. The 9-item shared decision making questionnaire (SDM-Q-9): development and psychometric properties in a primary care sample. *Patient Educ Couns*. 2010;80:94-9.
- Barr PJ, Thompson R, Walsh T, Grande SW, Ozanne EM, Elwyn G. The psychometric properties of collaborate: a fast and frugal patient-reported measure of the shared decision-making process. *J Med Internet Res*. 2014;16:e2.
- Forcino RC, Barr PJ, O'Malley AJ, Arend R, Castaldo MG, Ozanne EM, et al. Using collaborate, a brief patient-reported measure of shared decision making: results from three clinical settings in the United States. *Heal Expect*. 2018;21:82-9.
- Agency for Healthcare Research & Quality. About CAHPS. Retrieved April 24, 2018, from: <https://www.ahrq.gov/cahps/about-cahps/index.html>.
- Agency for Healthcare Research and Quality. Clinician & Group Survey and Instructions. Retrieved September 30, 2016, from: <http://www.ahrq.gov/cahps/surveys-guidance/cg/instructions/index.html>.
- Adams J. The Reliability of Provider Profiling: A Tutorial. RAND Corporation. 2009. Epub ahead of print 2009. doi:10.7249/TR653.
- Scholle SH, Roski J, Adams JL, Dunn DL, Kerr EA, Dugan DP, Jensen RE. Benchmarking physician performance: reliability of individual and composite measures. *Am J Manag Care*. 2008; 14: 833-8.
- Snijders TAB, Bosker RJ. *Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage Publications; 1999.
- Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. *J Gen Intern Med*. 2012;27:1361-7.
- Elwyn G, Durand MA, Song J, Aarts J, Barr PJ, Berger Z, et al. A three-talk model for shared decision making: multistage consultation process. *BMJ*. 2017;359:j4891.
- Wilson Vanvoorhis CR, Morgan BL. Understanding power and rules of thumb for determining sample sizes. *Tutor Quant Methods Psychol*. 2007;3:43-50.
- Hays RD, Berman LJ, Kanter MH, Kanter MH, Hugh M, Oglesby RR, et al. Evaluating the psychometric properties of the CAHPS patient-centered medical home survey. *Clin Ther*. 2014;36:689-696.e1.
- Clark EL. Spearman-brown formula applied to ratings of personality traits. *J Educ Psychol*. 1935;26:552-5.
- Paddison C, Elliott M, Parker R, Staetsky L, Lyratzopoulos G, John LC, Martin R. Should measures of patient experience in primary care be adjusted for case mix? Evidence from the english general practice patient survey. *BMJ Qual Saf*. 2012; 21:634-40.
- Raleigh V, Sizmur S, Tian Y, Thompson J. Impact of case-mix on comparisons of patient-reported experience in NHS acute hospital trusts in england. *J Health Serv Res Policy*. 2015;20: 92-9.

Author Biographies

Rachel C Forcino is research project coordinator at The Dartmouth Institute for Health Policy and Clinical Practice.

Marcus Thygeson is chief health officer at Bind Benefits.

A James O'Malley is professor at The Dartmouth Institute for Health Policy and Clinical Practice and the Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth.

Marjan J Meinders is senior researcher at the Scientific Center for Quality of Healthcare, Radboud University Medical Center.

Gert P Westert is professor and head of the Scientific Center for Quality of Healthcare, Radboud University Medical Center.

Glyn Elwyn is professor at The Dartmouth Institute for Health Policy and Clinical Practice.