

Improving Patients' Choice of Clinician by Including Roll-up Measures in Public Healthcare Quality Reports: an Online Experiment

Jennifer L. Cerully, Ph.D.¹, Andrew M. Parker, Ph.D.¹, Lise Rybowski, M.B.A.², Mark Schlesinger, Ph.D.³, Dale Shaller, M.P.A.⁴, Rachel Grob, Ph.D.⁵, Melissa L. Finucane, Ph.D.¹, and Steven C. Martino, Ph.D.¹

¹RAND Corporation, Pittsburgh, PA, USA; ²Severyn Group, Ashburn, VA, USA; ³Yale University, New Haven, CT, USA; ⁴Shaller Consulting Group, Chesterfield, MI, USA; ⁵University of Wisconsin–Madison, Madison, WI, USA.

BACKGROUND: Public reports on healthcare quality typically include complex data. To lower the cognitive burden of interpreting these data, some report designers create summary, or roll-up, measures combining multiple indicators of quality into one score. Little is known about how the availability of roll-ups affects clinician choice.

OBJECTIVE: To determine how presenting quality scores at different levels of aggregation affects patients' clinician choices.

DESIGN: We conducted a simulated clinician-choice experiment, randomizing participants to three versions of a public reporting website and comparing their clinician choices. One version aggregated all clinician-level quality measures into roll-ups, the second provided disaggregated (drill-down) scores only, and the third offered both roll-ups and drill-downs.

PARTICIPANTS: Five hundred fifty panelists drawn from a probability-based Internet panel.

MAIN MEASURES: We assessed the amount of effort participants exerted by tracking the length of time spent on the website and the number of concrete actions taken on the website (e.g., clicking items). We evaluated decision quality by measuring whether participants selected a clinician who performed more poorly than others and incongruence between participants' stated preferences for dimensions of quality and their chosen clinician's performance on those dimensions.

KEY RESULTS: Participants seeing drill-downs alone (mean = 14.9) or with roll-ups (mean = 19.2) took more actions than those who saw roll-ups alone (mean = 10.5) ($ps < 0.05$). However, participants seeing only drill-downs made poorer choices than those who saw roll-ups alone or with drill-downs. More participants seeing drill-downs chose a clinician who was outperformed (36.3% versus 23.4% [roll-up] and 25.6% [drill-down + roll-up], $ps < 0.05$) and made choices incongruent with stated preferences (51.2% versus 45.6% [roll-up] and 47.5% [drill-down + roll-up], $ps < 0.05$). The distinction between roll-up and drill-down was somewhat stronger for sicker participants.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11606-018-4725-y>) contains supplementary material, which is available to authorized users.

Received February 14, 2018

Revised August 28, 2018

Accepted October 4, 2018

Published online November 16, 2018

CONCLUSIONS: Our results suggest that roll-ups in healthcare quality reports, alone or as a complement to drill-downs, can help patients make better decisions for themselves.

KEY WORDS: healthcare quality; patient experience; patient satisfaction; composite measure; summary score.

J Gen Intern Med 34(2):243–9

DOI: 10.1007/s11606-018-4725-y

© Society of General Internal Medicine 2018

INTRODUCTION

Background

A primary objective of public reports on healthcare quality is to provide comparative information that patients can use to make well-informed choices among providers and health plans.¹ One critical challenge facing report designers is that this information typically consists of a large array of quality measures unfamiliar to most patients, as well as data that can be statistically complex and hard to decipher. To ease the cognitive task of interpreting and integrating many different pieces of information, report designers often employ multiple strategies, including the use of symbols to convey relative performance, the use of plain language to make measures understandable and salient, and the use of summary, or roll-up, measures that combine multiple, often disparate indicators of healthcare quality into a single score.^{1–9}

At the highest level of aggregation, a roll-up quality measure can combine indicators of patient experience, clinical processes, and patient outcomes. One prominent example of this kind of measure is the Centers for Medicare & Medicaid Services (CMS)' Five-Star Quality Rating System, which assigns overall ratings of one to five stars to different providers and health plans based on their performance across various domains of quality.¹⁰ A roll-up measure could also represent an organization's performance across multiple indicators of a single dimension of quality.¹¹ For example, for medical groups, the California Office of the Patient Advocate reports a score for "Patients Rate Overall Experience" that combines five patient experience measures.

The use of roll-up measures to reduce the cognitive burden imposed by quality reports is predicated on three assumptions: (1) well-constructed roll-up measures can accurately summarize performance across different dimensions of care,¹² (2) roll-up measures can capture enough performance variation for patients to choose wisely among alternatives, and (3) weighing fewer attributes (e.g., quality measures) results in better decisions.^{6, 13}

Not all stakeholder groups (including patients, clinicians, hospitals, health plans, and developers of healthcare quality reports) view the provision of roll-up scores in healthcare quality reports for patients as a positive strategy. Critics commonly express concerns about the potential of roll-ups to negatively affect decision making, obscuring important nuances of quality, the sometimes-hidden assumptions about the relative importance of component measures, and the lack of evidence-based methods for calculating scores.^{11, 14–16} Additionally, some report developers raise concerns about whether display strategies align with patients' preferences.^{11, 14–20}

Objective

The public reporting community is divided on the use of roll-ups, but little is known about patients' understanding and use of such scores when choosing clinicians.¹¹ To address that gap, we conducted an experiment to determine whether presenting quality scores at different levels of aggregation affects patients' clinician choices. A realistic, interactive website presented participants with information about the quality of a set of fictional clinicians. This information included scores for an array of quality measures as well as patients' comments about clinicians. The site allowed participants to explore the performance of different clinicians and make a hypothetical selection. We used this experimental platform to explore how the choice of clinician varies based on exposure to roll-up scores, disaggregated drill-down scores (i.e., the component measures used to create a roll-up measure), or both. The third arm was included because several report developers participating in a previous study indicated that they would most likely present both roll-up and drill-down scores rather than just one.¹¹

Because of the complexity and volume of healthcare quality data, we anticipated that organizing and presenting data in ways that are easier for participants to understand (i.e., roll-up scores) would result in a lower cognitive load, leading to an easier decision process, and better choices of clinicians. Specifically, we hypothesized that relative to those presented with drill-down scores only, participants presented with roll-up scores would exert less effort in the choice process, be more likely to choose the best-performing clinician (according to quantified performance metrics), and make fewer preference-incongruent choices. We entertained two competing hypotheses for participants seeing both roll-up and drill-down scores. On the one hand, because those who saw both roll-up and drill-down scores could view quality data at whatever level of

aggregation aligns with personal information needs and preferences, they might behave similarly to those in the roll-up arm. On the other hand, participants receiving both roll-up and drill-down scores might have too much information to process, putting them at risk of confusion and poorer choices.

Finally, we explored whether the effects of presenting roll-up scores differed across sicker and healthier participants. Because sicker participants have more experience interacting with clinicians, they might value and be able to use detailed information about clinician performance more than healthier participants do. Alternatively, sicker participants who are taxed by their poor health may find detailed information to be especially overwhelming. If so, the availability of drill-down data might be detrimental to the choices of sicker versus healthier participants.

METHODS

Data were collected from May to July 2015 as part of a larger study of clinician choice.²¹ We focus here on three experimental arms that varied the types of scores provided to participants. This study was approved by the first author's institutional review board. All participants underwent an informed consent procedure.

Participants

Participants were drawn from the GfK probability-based Internet Knowledge Panel, which is designed to be representative of the US population.²² Fifty-two percent of panelists were invited to participate in the study accepted, resulting in 550 participants across the three arms.

Design and Procedure

Participants completed an initial survey that included questions about their real-life healthcare experiences and how they choose a clinician. After a week, 85% of the respondents who completed the initial survey returned to the study and were directed to an experimental website called SelectMD that had information about a fictitious set of 12 clinicians (see Fig. 1). After being randomly assigned to one of three experimental arms, participants were asked to use the information on the website to select the clinician they thought would be best for their healthcare needs, treating the choice as carefully as if it were a real one (see a related methodological report,²¹ or the [online appendix](#) for more on the site). After selecting a clinician, participants completed a second survey about the choice process. The SelectMD website collected data on participants' website use while they were engaged in the choice task. The analyses presented in this paper combine objective measures of website usage and participants' survey responses. Most data were based on self-reports or behavior within the study. Some socio-demographic variables (gender, age, education, and

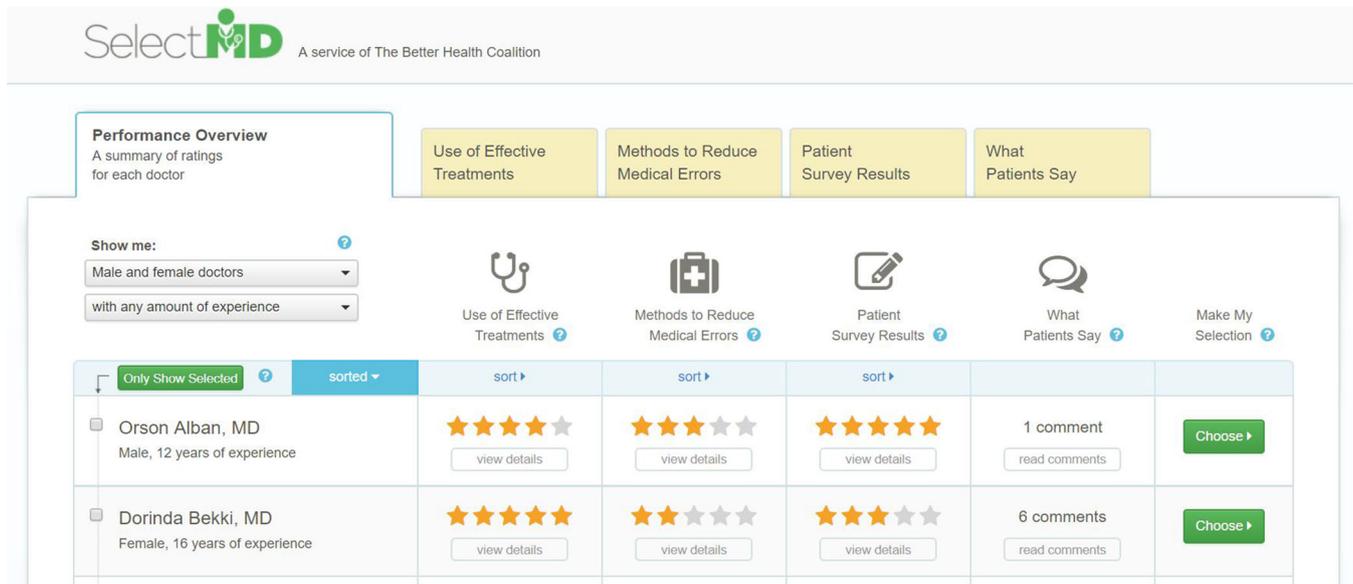


Figure 1 Sample screenshot from the SelectMD experimental website.

race) were maintained by GfK as part of panel administration and provided to us.

The SelectMD website allowed participants to see three categories of quantitative performance metrics: patient experience survey scores, clinical quality scores (i.e., indicators of the extent to which the clinician-delivered care aligned with best practice), and patient safety scores (i.e., indicators of the clinician's adoption of protocols that enhance patient safety). Scores were presented as one-to-five star ratings. In addition, all participants could view comments about clinicians and their staff; comments did not differ across the experimental arms. These comments mimicked actual patient comments about clinicians; the content was drawn from real comments found on clinician quality websites and comments elicited in a companion study.²³

The three experimental arms varied the presentation of performance scores. In the drill-down arm, participants saw scores for four measures in each of the three categories, for a total of 12 quality scores (e.g., for patient experience, scores indicated how well clinicians communicate with their patients, patients' ability to get timely appointments and information, the courtesy and helpfulness of office staff, and how well clinicians attend to their patients' mental or emotional health). In the roll-up arm, participants saw only rolled-up scores for each of the three categories of metrics (patient experience, clinical quality, and patient safety); they did not see the individual measures combined to form each roll-up score. In the drill-down plus roll-up arm, participants could see roll-up scores for each category as well as the individual measures used to form the roll-up scores.

Main Measures

Decision Quality. We measured decision quality in two ways. First, some presented clinicians were objectively worse than other

presented clinicians, in the sense that they were at least as bad on all dimensions (i.e., the three categories of metrics) and worse on at least one dimension (this is also known as a "dominated" option²⁴). We assessed whether a participant chose a suboptimal clinician.

Second, we constructed a measure, labeled preference incongruence, that captured the level of disagreement between how important a participant said each of the dimensions was in a preferred clinician and the characteristics of the selected clinician. For each of the three dimensions reported on the website, participants rated whether each dimension "matters a lot," "matters some," or "does not matter much." Participants were labeled as having made a preference-incongruent choice if they identified one of the performance dimensions as being among their top characteristics of a preferred clinician (i.e., they reported that the dimension "matters a lot") but did not actually select a clinician on the website who was highly rated on that dimension (i.e., had a five-star rating). For example, a participant who indicated that safety matters a lot and chose a clinician with three stars on safety would have made a preference-incongruent choice on that dimension. This preference incongruence measure was constructed using the preferences expressed on the post-choice survey to allow for learning from the website.²⁵

Level of Effort. We used two objective measures derived from tracking data collected while participants used the SelectMD website to determine level of effort: the length of time spent on the website and the number of concrete actions taken (e.g., clicking on items, hovering over a pop-up element).

Assessing Health Status. As part of the initial survey, participants were asked if they had ever been treated for a serious or life-threatening condition. They were also asked if

Table 1 Sample Characteristics by Experimental Arm

Participant characteristics	Experimental arm			Chi-square test for cross-arm differences	Percent of adults in US population (2010)
	Drill-down + roll-up	Roll-up only	Drill-down only		
	<i>n</i> = 176 (%)	<i>n</i> = 184 (%)	<i>n</i> = 190 (%)		
Socio-demographics*					
Age					
18–29	13.6	10.9	11.1	<i>p</i> = 0.75	23.1†
30–44	22.2	26.1	23.7		24.1
45–60	30.1	31.0	26.3		25.1
> 60	34.1	32.1	38.9		27.7
Race/ethnicity					
White	78.4	74.5	78.9	<i>p</i> = 0.17	69.6
African American	4.5	8.7	6.8		11.2
Latino	3.4	8.2	5.8		12.7
Other	13.6	8.7	8.4		6.5
Education					
High school or less	31.3	33.2	35.8	<i>p</i> = 0.92	44.1
Some college	27.8	27.7	26.8		25.9
College graduate	40.9	39.1	37.4		29.9
Health status and utilization‡					
Chronic health problems§	52.8	44.6	45.8	<i>p</i> = 0.24	45.7
Self-reported health					
Fair/poor	14.8	10.3	13.7	<i>p</i> = 0.18	12.9¶
Good	28.4	37.0	38.9		26.7
Very good/excellent	56.8	52.7	47.4		60.3
Prior use of doctor quality websites					
Visited a website in the past 12 months to learn about doctor quality#	41.1	41.4	45.6	<i>p</i> = 0.74	32–40**

*These data are maintained by GfK as part of panel administration and were provided to the study team along with the final study dataset

†From US Census, Statistical Abstract of the United States: 2012²⁶

‡Participants reported these data in response to questions included as part of the study

§Chronic health condition refers to those who reported having been treated for a serious or life-threatening condition and/or having a long-term medical condition requiring medical monitoring or treatment

||From Ward et al.²⁷

¶From Blackwell et al.²⁸

#As part of the post choice survey, participants were asked if in the past 12 months, they had visited an Internet website to learn specifically about the quality of doctors in their area. The percentages are of those participants who selected “yes” (with other response options being “no” and “not sure”)

**From an Associated Press-NORC Center for Public Affairs study in which participants indicated whether in the past 12 months, they had seen or heard information about the quality of doctors or other healthcare providers from ratings websites, community or advocacy group websites, or government websites.²⁹

they had a long-term medical condition that required medical monitoring or treatment. Participants who responded “yes” to either question were classified as “sicker;” all others were coded as “healthier.”

Statistical Analysis

We examined sample characteristics and tested the success of random assignment across the three arms. Then, we tested for pairwise differences among the three arms on each of the dependent measures, using chi-square tests of independence and independent-sample *t* tests. Finally, we broke the sample into healthier and sicker subsamples, again testing for differences across arms on each of the dependent measures.

RESULTS

As shown in Table 1, participants did not differ across experimental arms in terms of age, race/ethnicity, education, chronic disease, self-reported health status, and self-reports of having

visited a doctor quality website, indicating that randomization to experimental arm was effective. Because there were no significant differences in patient characteristics across arms, these characteristics were not included as covariates in the models presented here.¹ Compared to the general US population, our sample skewed older, whiter, and more educated. Our sample was comparable to the US population in terms of the proportion experiencing chronic health conditions but reported slightly poorer health.

Effect of Provision of Roll-up Scores on Level of Effort

Participants who saw drill-down scores (mean = 14.9, SD = 19.0; Table 2) or drill-down scores and roll-up scores together (mean = 19.2, SD = 28.5) took significantly more actions while navigating the site relative to those who saw roll-up

¹As a robustness check, we conducted regression analyses using arm as a predictor and age, race/ethnicity, education, chronic condition, self-reported health status, and self-reports of visiting a doctor quality website as control variables. These models generally yielded the same findings on the key outcomes, so we report the simpler models.

Table 2 Effects of Quality Data Presentation on Decision Quality

Categories of quality	Measure (scale)	Mean (SD) by experimental arm		
		Drill-down plus roll-up (n = 176)	Roll-up (n = 184)	Drill-down (n = 190)
Amount of effort	Time deliberating (number of seconds)	437.6 (482.9)	373.9 (321.7)	450.4 (724.6)
	Active exploring of website (number of actions)	19.2 ^a (28.5)	10.5 ^{ab} (9.5)	14.9 ^b (19.0)
Quality of decision	Proportion of times clinician choice did not match with stated preferences (percentage)	47.5 ^a (14.7)	45.6 ^b (16.0)	51.2 ^{ab} (14.5)
	Choice of a suboptimal clinician (percentage)	25.6 ^a	23.4 ^b	36.3 ^{ab}

Means/percentages within the same row sharing the same superscript are significantly different ($p < 0.05$) from one another

scores (mean = 10.5, SD = 9.5).

Effect of Provision of Roll-up Scores on Whether the Best Clinician Is Chosen

A significantly greater proportion of participants in the drill-down arm chose a suboptimal clinician (36.3%, Table 2) relative to participants in the roll-up (23.4%) and drill-down plus roll-up (25.6%) arms.

Participants in the drill-down arm made more preference-incongruent choices (51.2%, Table 2), in contrast to participants in the roll-up arm (45.6%) and participants in the drill-down plus roll-up arm (47.5%).

Thus, on both measures of decision quality, more participants in the drill-down arm made poor quality choices than did participants in the other arms.

Differences Based on Health Status

We examined the level of effort and decision quality among healthier and sicker participants (Table 3). For neither group did time spent deliberating vary across the arms, and both groups took more actions in the drill-down plus roll-up arm than in the roll-up arm (healthier participants also took more actions in the drill-down plus roll-up arm than in the drill-down arm).²

²A regression analysis using arm as a predictor and age, race, education, having a chronic condition, self-reported health, and self-reports of visiting doctor quality websites as control variables resulted in changes to findings related to the number of actions taken. Two significant differences crossed the 0.05 threshold after controlling for covariates, becoming non-significant. The difference between drill-down plus roll-up and drill-down only went from significant ($p = 0.049$) to non-significant ($p = 0.062$), as did the difference between roll-up only and drill-down only (going from $p = 0.028$ to $p = 0.056$).

The distinction between roll-up and drill-down on decision quality was somewhat stronger for sicker participants than healthier ones. Specifically, among sicker participants, a greater proportion chose a suboptimal clinician or made preference-incongruent choices when viewing drill-down scores relative to those who saw roll-up scores (or drill-down plus roll-up scores for preference incongruence). Among healthier participants, more participants in the drill-down arm selected a suboptimal clinician than did those in the drill-down plus roll-up arm, and those who viewed only drill-down scores made more preference-incongruent choices relative to those who saw roll-up scores.

DISCUSSION

In summary, participants in this study who saw quality information only in the form of drill-down scores tended to engage in more effortful consideration of the data they were provided (as indicated by taking more actions on the website) but made worse clinician choices than did those who had access to roll-up scores. This was true whether decision quality was measured by preference incongruence or the selection of dominated alternatives. Adding drill-down scores to roll-up scores increased effort but did not appear to harm decision quality. These findings suggest that it is advantageous to include roll-up scores in reports on healthcare quality with or without accompanying drill-down information.

The differences between healthier and sicker patients in their use of quality metrics may be useful for report developers. Because sicker participants likely have more frequent encounters with clinicians than do healthy participants, the process of choosing a clinician is especially salient for them. Information presentation appeared to affect the amount of effort healthier participants exerted, yet had relatively little effect on decision quality. Conversely, while sicker participants also differed in effort across experimental arm (although less so than healthier participants), the differences in decision quality were more profound, with those seeing only drill-down information performing the worst. This finding suggests that providing roll-up information (with or without accompanying drill-down information) may be particularly important for those most involved with the healthcare system. As posited earlier, sicker patients may be taxed by their poor health and benefit from the reduced cognitive load provided by roll-up scores. However, this explanation cannot be tested here, and further studies should determine the robustness of this difference.

The study is subject to several limitations. First, while the website was designed to be as realistic as possible, participants engaged in a hypothetical clinician choice, not a real one. Despite the hypothetical nature of the task, participants spent about 6 min to 7.5 min choosing a clinician, suggesting they engaged in the task. Though the study was not designed to estimate rates of poor decision making in the general population, the rates of suboptimal choices in this study (23% to

Table 3 Effects of Quality Data Presentation on Decision Quality for Healthier and Sicker Participants

Categories of quality	Measure (scale)	Mean (SD) experimental arm					
		Drill-down plus roll-up	Roll-up	Drill-down	Drill-down plus roll-up	Roll-up	Drill-down
		Healthier participants			Sicker participants		
		<i>n</i> = 83	<i>n</i> = 102	<i>n</i> = 103	<i>n</i> = 93	<i>n</i> = 82	<i>n</i> = 87
Amount of effort	Time deliberating (number of seconds)	466.4 (553.5)	365.1 (336.4)	378.4 (490.5)	411.4 (409.5)	384.8 (304.3)	534.9 (922.9)
	Active exploring of website (number of actions)	21.0 ^{ab} (31.9)	10.0 ^a (8.8)	14.0 ^b (17.2)	17.5 ^a (25.1)	11.1 ^a (10.4)	16.0 (20.9)
Quality of decision	Proportion of times clinician choice did not match with stated preferences (percentage)	48.0 (13.5)	45.1 ^a (16.8)	50.3 ^a (15.4)	47.3 ^a (15.8)	46.1 ^b (14.9)	52.3 ^{ab} (13.3)
	Choice of a suboptimal clinician (percentage)	20.5 ^a	25.5	34.0 ^a	30.1	20.7 ^a	39.1 ^a

Healthier participants are those who reported having neither a serious or life-threatening condition nor a long-term medical condition requiring medical monitoring or treatment. Sicker participants were those who reported having been treated for a serious or life-threatening condition and/or having a long-term medical condition requiring medical monitoring or treatment. Means/percentages within the same row sharing the same superscript are significantly different ($p < 0.05$) from one another

36%, depending on arm) are lower than those found in other studies of real-world health decision making.^{30, 31} For example, from 2007 to 2010, traditional Medicare was a dominated option relative to Medicare Advantage, but less than a quarter of beneficiary-selected Medicare Advantage.³⁰ Second, though we attempted to capture the most relevant constructs related to physician choice, unmeasured factors could have affected responses. Randomization to experimental arms should have minimized the chance that variation in unmeasured constructs would vary systematically across arms. Third, the study focused on the choice of clinician only. We did not test other healthcare choice scenarios (e.g., selecting hospitals, health plans), but we would not predict that the effect of providing roll-up scores would differ substantially.

This study increases knowledge of how patients' clinician choices are influenced by the level of aggregation and presentation of healthcare quality data. Some questions about roll-up scores remain unaddressed, including how clinicians or health plans respond to roll-up measures or use them to guide quality improvement initiatives^{32, 33} and determining best methods of roll-up score calculation.

CONCLUSIONS

The results of this study suggest the value of presenting roll-up scores in healthcare quality reports for patients. Developers of such reports may want to consider summarizing performance, where possible, both to reduce cognitive load and to improve decision quality. Because participants in the roll-up only and drill-down plus roll-up arms did not significantly differ in their likelihood of making errors, it appears that providing drill-down scores does not necessarily hurt decision making. Instead, it seems that roll-up scores (whether presented alone or as a complement to drill-down scores) can potentially improve decision making. Previous research suggests that report

sponsors and national organizations involved in public reporting may be more comfortable with offering both roll-up and drill-down measures to accommodate individuals with different information needs.¹¹ Though the reporting of roll-up scores continues to pose practical issues (e.g., the development of scientific methods for calculating fair and reliable scores¹¹), this study is the first to empirically demonstrate that the provision of roll-up scores can increase the proportion of patients choosing better performing clinicians.

Corresponding Author: Jennifer L. Cerully, Ph.D.; RAND Corporation, 4570 Fifth Avenue, Suite 600, Pittsburgh, PA, USA (e-mail: jcerully@rand.org).

Funding This paper was supported by two cooperative agreements (2U18HS016980 and 1U18HS016978) from AHRQ to RAND Corporation and Yale University, respectively.

Compliance with Ethical Standards:

Conflict of Interest: The authors declare that they do not have a conflict of interest.

Prior Presentations: This research has not been presented previously.

REFERENCES

1. Findlay SD. Consumers' interest in provider ratings grows, and improved report cards and other steps could accelerate their use. *Health Aff.* 2016;35(4):688–96.
2. Martino SC, Kanouse DE, Miranda DJ, Elliott MN. Can a more user-friendly Medicare Plan Finder improve consumers' selection of Medicare plans?. *Health Serv Res.* 2017;52(5):1749–71.
3. Sinaiko AD, Eastman D, Rosenthal MB. How report cards on physicians, physician groups, and hospitals can have greater impact on consumer choices. *Health Aff.* 2012;31(3):602–11.
4. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. *Annu Rev Public Health.* 2003;24:413–33.

5. **Kurtzman ET, Greene J.** Effective presentation of health care performance information for consumer decision making: a systematic review. *Patient Educ Couns.* 2016;99(1):36–43.
6. **Schlesinger M, Kanouse DE, Martino SC, Shaller D, Rybowski L.** Complexity, public reporting, and choice of doctors: a look inside the blackest box of consumer behavior. *Med Care Res Rev.* 2014;71(5 Suppl):38s–64s.
7. **Schlesinger M, Kanouse DE, Rybowski L, Martino SC, Shaller D.** Consumer response to patient experience measures in complex information environments. *Med Care.* 2012;50(Suppl):s56–64.
8. **Gebele C, Tscheulin DK, Lindenmeier J, Drevs F, Seemann AK.** Applying the concept of consumer confusion to healthcare: development and validation of a patient confusion model. *Health Services Management Research.* 2014;27(1–2):10–21.
9. **Shi Y, Scanlon DP, Bhandari N, Christianson JB.** Is anyone paying attention to physician report cards? The impact of increased availability on consumers' awareness and use of physician quality information. *Health Serv Res.* 2017;52(4):1570–89.
10. Centers for Medicare and Medicaid Services. Hospital Compare star ratings fact sheet. Available at: <https://www.cms.gov/newsroom/fact-sheets/hospital-compare-star-ratings-fact-sheet>. Accessed 10 September 2018.
11. **Cerully JL, Martino SC, Rybowski L, Finucane ML, Grob R, Parker AM, Schlesinger M, Shaller D, Martsoff GR.** Using “roll-up” measures in health care quality reports: perspectives of report sponsors and national alliances. *American Journal of Managed Care.* 2017;23(6):e202–207.
12. **Zaslavsky AM, Shaul JA, Zaborski LB, Cioffi MJ, Cleary PD.** Combining health plan performance indicators into simpler composite measures. *Health Care Financing Review.* 2002;23(4):101–15.
13. **Schwartz B.** *The Paradox of Choice: Why More is Less.* New York, NY, US: Harper Collins Publishers; 2004.
14. **Martsoff GR, Scanlon DP, Christianson JB.** Multistakeholder perspectives on composite measures of ambulatory care quality: a qualitative descriptive study. *Med Care Res Rev.* 2013;70(4):434–48.
15. Association of American Medical Colleges. Guiding principles for public reporting of provider performance. Available at: <https://www.aamc.org/download/370236/data/guidingprinciplesforpublicreporting.pdf>. Accessed 10 September 2018.
16. **Thompson A.** RE: Enhancements of the Overall Hospital Quality Star Rating, August 2017. Available at: <https://www.aha.org/system/files/advocacy-issues/letter/2017/170925-let-thompson-cms-star-ratings.pdf>. Accessed 10 September 2018.
17. **Romano P, Hussey P, Ritley D.** Selecting quality and resource use measures: a decision guide for community quality collaboratives. Available at: <https://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/perfmeasguide/index.html>. Accessed 10 September 2018.
18. **Boyce T, Dixon A, Fasolo B, Reutskaja E.** Choosing a high-quality hospital: the role of nudges, scorecard design and information. Available at: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Choosing-high-quality-hospital-role-report-Tammy-Boyce-Anna-Dixon-November2010.pdf. Accessed 10 September 2018.
19. **Orlowski JM.** Re: AAMC comments on the measure selection for Hospital Compare star ratings TEP report. Available at: <https://www.aamc.org/download/425936/data/aamccommentletteroncmsstarratingstep.pdf>. Accessed 10 September 2018.
20. **Nickels TP.** RE: CMS-1677-P, Medicare Program; Hospital Inpatient Prospective Payment Systems (PPS) for Acute Care Hospitals and the Long-Term Care Hospital PPS and Proposed Policy Changes and FY 2018 Rates; Quality Reporting Requirements for Specific Providers; Medicare and Medicaid Electronic Health Record (EHR) Incentive Program Requirements for Eligible Hospitals, Critical Access Hospitals, and Eligible Professionals; Provider-Based Status of Indian Health Service and Tribal Facilities and Organizations; Costs Reporting and Provider Requirements; Agreement Termination Notices; Proposed Rule (Vol. 82, No. 81) April 28, 2017. Available at: <https://www.aha.org/system/files/advocacy-issues/letter/2017/170613-let-nickels-verma-cms-flexibility-efficiencies.pdf>. Accessed 10 September 2018.
21. **Cerully JL, Parker AM, Schlesinger M, Martino SC, Shaller D, Rybowski L, Grob R, Finucane ML.** The Select MD 2.0 Provider Choice Experiment: Methodological Overview. Santa Monica, CA: RAND Corporation; 2016. Available at: https://www.rand.org/pubs/working_papers/WR1152.html. Accessed 10 September 2018.
22. GfK. Knowledge Panel® design summary. Available at: [http://www.knowledgenetworks.com/knpanel/docs/KnowledgePanel\(R\)-Design-Summary-Description.pdf](http://www.knowledgenetworks.com/knpanel/docs/KnowledgePanel(R)-Design-Summary-Description.pdf). Accessed 10 September 2018.
23. **Grob R, Schlesinger M, Parker AM, Shaller D, Barre LR, Martino SC, Finucane ML, Rybowski L, Cerully JL.** Breaking narrative ground: innovative methods for rigorously eliciting and assessing patient narratives. *Health Serv Res.* 2016;51(Suppl 2):1248–72.
24. **Montgomery H.** Decision rules and the search for a dominance structure: Towards a process model of decision making. *Advances in Psychology.* 1983;14:343–69.
25. **Fowler FJ Jr, Levin CA, Sepucha KR.** Informing and involving patients to improve the quality of medical decisions. *Health Aff.* 2011;30(4):699–706.
26. United States Census Bureau. Statistical abstract of the United States: 2012. Available at: <https://www.census.gov/library/publications/2011/compendia/statab/131ed.html>. Accessed 10 September 2018.
27. **Ward BW, Schiller JS, Goodman RA.** Multiple chronic conditions among US adults: a 2012 update. *Preventing Chronic Disease.* 2014;11.
28. **Blackwell DL, Lucas JW, Clarke TC.** Summary health statistics for U.S. adults: national health interview survey, 2012. *Vital and Health Statistics, Series 10, Number 260.* 2014.
29. Associated Press – NORC Center for Public Affairs. Finding quality doctors: how Americans evaluate provider quality in the United States. Available at: <http://www.apnorc.org/projects/Pages/finding-quality-doctors-how-americans-evaluate-provider-quality-in-the-united-states.aspx>. Accessed 10 September 2018.
30. **Darby C, Crofton C, Clancy CM.** Consumer Assessment of Health Providers and Systems (CAHPS): evolving to meet stakeholder needs. *Am J Med Qual.* 2006;21(2):144–7.
31. Agency for Healthcare Research and Quality. Research on improving patient experience. Available at: <https://www.ahrq.gov/cahps/quality-improvement/reports-and-case-studies/index.html>. Accessed 10 September 2018.
32. **Afendulis CC, Sinaiko AD, Frank RG.** Dominated choices and Medicare Advantage enrollment. *J Health Econ.* 2015;119:72–83.
33. **Sinaiko AD, Hirth RA.** Consumers, health insurance and dominated choices. *J Health Econ.* 2011;30:450–457.