VIEWPOINT

Considerations of Bias and Reliability in Publicly Reported Physician Ratings

Leah M. Marcotte, MD¹ , Rachel B. Issaka, MD, MAS^{2,3}, and Nidhi Agrawal, PhD⁴

¹Division of General Internal Medicine, Department of Medicine, University of Washington School of Medicine, Seattle, WA, USA; ²Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; ³Division of Gastroenterology, Department of Medicine, University of Washington School of Medicine, Seattle, WA, USA; ⁴Department of Marketing and International Business, Foster School of Business, University of Washington, Seattle, WA, USA.

J Gen Intern Med 36(12):3857–8

DOI: 10.1007/s11606-021-06898-z

© Society of General Internal Medicine 2021

Many health systems have opted to publicly report physician ratings and patient comments alongside physician profiles on their websites. While this is meant to engage consumers to make informed decisions in choosing physicians, these ratings are subject to biases and poor statistical reliability that may disproportionately disadvantage certain groups of physicians.¹ Health systems should be aware of and evaluate disparities arising from biases prior to publicly sharing these data.

BIASES IN RATINGS

Cognitive biases such as implicit bias—in which stereotypes of specific groups unconsciously drive behavior and decision-making—may impact patient ratings of physicians. Implicit bias based on race, ethnicity, and gender has been demonstrated to impact ratings of individuals and businesses outside of health care. For example, projects posted by Black people on crowdfunding platforms are seen as lower quality and are less likely to be funded.² Additionally, businesses in Black neighborhoods in the USA receive fewer and lower ratings on Yelp.³ In academia, ratings of professors by students are subject to gender and ethnic biases. And a recent evaluation of direct-to-consumer telehealth services, for instance, showed lower ratings of non-white physicians.⁴ While the data in health care is scant, implicit bias likely impacts physician ratings.

Other types of biases may compound the effects of implicit bias. Information biases—i.e., bias due to measurement error—may also impact ratings. For example, a patient may fill out a survey intended for their general internist, but instead

rate their oncologist. Or a patient may misidentify their physician, inadvertently rating another physician of the same race or ethnicity in practices with few underrepresented minority physicians.⁵ These examples come from anecdotal experience of the authors and certainly more research is needed to understand the frequency of information bias in physician ratings. It is plausible, however, that information bias may compound or result from implicit bias.

Selection bias—i.e., bias that occurs when a study population (in this case survey respondents) is not representative of the general population—is well established in ratings outside of health care and likely also plays a role in physician ratings. Physician ratings are on average high and are considered a “top-box” measure. In specialties where physicians see patients longitudinally, patients will generally stay with physicians they favor and leave physicians who are not a good fit. Additionally, physician-patient relationships evolve, and patients may be more likely to rate physicians and to rate them more highly after building a trusting relationship. Thus, selection bias may specifically favor more senior physicians and adversely impact early career physicians or those who are new to a health care system.

For physicians with intersecting identities (e.g., an early career Black woman), compounding biases may lead to harmful ratings with immeasurable downstream consequences. To minimize the impact of bias in physician ratings and any potential harm, health systems that choose to use publicly available physician rating systems must routinely evaluate these platforms and any data generated with equity in mind.

STATISTICAL RELIABILITY

Health systems receive relatively few ratings for each physician, raising concerns regarding statistical reliability. Reliability can be defined as the reproducibility of a measure and increases with the number of measurements. Generally, reliability of 90% or above is desirable to ensure that the measure is adequately capturing data. The most commonly reported ratings of physicians are based on results from the Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey, which provides validated metrics of patient experiences including ratings of clinicians’ communication, respect,

Leah M. Marcotte and Rachel B. Issaka contributed equally to this work. Received February 6, 2021

Accepted May 3, 2021

Published online May 19, 2021

knowledge of medical history, and time spent. To reach adequate statistical reliability for a primary care physician, at least 138 and ideally 255 surveys to account for patient characteristics are required in a year.⁶ Yet, many systems have a threshold to publicly report physician ratings at 30 surveys or even less. Health systems outsource collecting provider ratings to companies that ultimately determine the cost of each completed survey. The challenge for health systems, especially academic medical centers where physicians commonly spend a fraction of their time on patient care, is that it becomes difficult and expensive to reach 255 surveys for each provider in a given year.

UNINTENDED CONSEQUENCES IN PUBLIC REPORTING OF PHYSICIAN RATINGS

Once ratings are publicly shared, physicians with lower ratings could face confirmation bias—i.e., bias that affects new experiences based on previously known information. To illustrate, a patient who chooses a physician with 4.9 stars may perceive their experience to be better than if they had no knowledge of ratings prior to their visit. Conversely, a patient who sees a physician who only has 3.9 stars may perceive their experience to be worse. Confirmation bias could affect a physician's ability to recruit new patients or evaluations by health care leadership and both phenomena ultimately reinforce high or low future ratings.

Publicly reporting physician ratings may create a greater emphasis within a health system for maintaining higher ratings. If physician hiring decisions, compensation, and promotions are directly or indirectly tied to ratings, then biased ratings could create systematic barriers for growth and inequities in professional and economic growth for physicians from backgrounds underrepresented in medicine.¹ Thus, the decision to publish physician ratings may ultimately be detrimental to health systems by undermining efforts to improve diversity, equity, and inclusion.

RECOMMENDATIONS TO HEALTH SYSTEMS

Health systems should systematically evaluate CAHPS and other data regarding physician ratings for biases. Academic institutions with access to researchers should lead the charge by reviewing these evaluations internally and partnering with non-academic institutions to do the same. Findings from these analyses that demonstrate the biases we outlined should be published to develop an evidence base for provider ratings and

contribute to growing literature on how to promote equity within health care systems' workforce.⁷

Instead of publishing individual provider ratings, health systems could instead provide average patient experience ratings at the clinic level. This could address poor reliability, effects of implicit bias of individuals, and confirmation bias when rating individual physicians, yet still give consumers information about patient experience measures in a given clinic. Clinic-level ratings would also better align with models of team-based care—a concept that is increasingly important in medicine. Of course, even at the clinic level, systems should still be sensitive to and evaluate potential biases, e.g., differences in ratings according to neighborhood. If found, the health system should work with Diversity, Equity, and Inclusion leadership to decide whether and how to report these findings. Ultimately, if bias is detected, the risk of publishing these data likely outweighs the potential benefit.

Acknowledgements: We thank Dr. Edwin Wong for providing valuable feedback to manuscript revisions. We thank Dr. Frederick Chen and Dr. Linnaea Schuttner for contributions to initial concept for this manuscript.

Corresponding Author: Leah M. Marcotte, MD; Division of General Internal Medicine, Department of Medicine, University of Washington School of Medicine, Seattle, WA, USA (e-mail: leahmar@uw.edu).

REFERENCES

1. **Poole KG Jr.** Patient-experience data and bias - what ratings don't tell us. *N Engl J Med.* 2019;380(9):801-803.
2. **Younkin P, Kuppawamy V.** The colorblind crowd? Founder race and performance in crowdfunding. *Management Science.* 2018; 64(7): 2973-3468.
3. **Perry AM, Rothwell J, Harshbarger D.** Five-star reviews, one-star profits: the devaluation of businesses in Black communities. The Metropolitan Policy Program at Brookings. 2020. Online. Available at: https://www.brookings.edu/wp-content/uploads/2020/02/2020.02_DevOfBizInBlackCommunities_Perry-Rothwell-Harshbarger-final.pdf. Accessed: 01/18/21.
4. **Martinez KA, Keenan K, Rastogi R, et al.** The association between physician race/ethnicity and patient satisfaction: an exploration in direct to consumer telemedicine. *J Gen Intern Med.* 2020; 35(9):2600-2606.
5. **Hughes BL, Camp NP, Gomez J, Natu VS, Grill-Spector K, Eberhardt JL.** Neural adaptation to faces reveals racial outgroup homogeneity effects in early perception. *Proc Natl Acad Sci U S A.* 2019 Jul 16;116(29):14532-14537.
6. **Fenton JJ, Jerant A, Kravitz RL, Bertakis KD, Tancredi DJ, Magnan EM, Franks P.** Reliability of physician-level measures of patient experience in primary care. *J Gen Intern Med.* 2017 Dec;32(12):1323-1329.
7. **Issaka RB.** Good for Us All. *JAMA.* 2020 Aug 11;324(6):556-557.

Publisher's Note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.