Author’s response to reviews

Title: How Are Medical Groups Identified as High-Performing? The Effect of Different Approaches to Classification of Performance

Authors:
Sangeeta Ahluwalia (sahluwal@rand.org)
Cheryl Damberg (damberg@rand.org)
Ann Haas (ahaas@rand.org)
Paul Shekelle (shekelle@rand.org)

Version: 1 Date: 01 Apr 2019

Author’s response to reviews:

Dear Dr. Zalm and the BMC Health Services Research Editorial Board:

Thank you for your review of our manuscript entitled “How Are Medical Groups Identified as High-Performing? The Effect of Different Approaches to Classification of Performance”. We greatly appreciate the reviewers’ thoughtful comments, and the opportunity to revise our manuscript accordingly. Responses to each of the comments are outlined below, and all edited or new text has been entered into the manuscript in the “Track Changes” format in MS Word.
Reviewer #1

1. There was limited background and the literature in relation to composite measures such as provided by Gaynor may be useful in strengthening the paper.

We now include a discussion of the advantages and disadvantages of composite measures (versus individual measures, both of which are included in the data we used for this study) in the Background section (Page 5, Lines 101-111). Though we were unable to find the specific reference cited by the reviewer, we have incorporated relevant context from Shwartz et al (2015) on the use of composite measures of performance.

2. I felt that the lack of risk adjustment (and lack of discussion regarding risk adjustment) was a limitation and if not possible given the data, should be identified as a limitation.

Thank you for raising this important point, particularly given recent discussion regarding the role of social risk factors in performance measurement (see for example, “Evaluation of the NQF Trial Period for Risk Adjustment for Social Risk Factors” by the National Quality Forum available at: https://www.qualityforum.org/Publications/2017/07/Social_Risk_Trial_Final_Report.aspx). We now raise this point in our Discussion section, Page 14, Lines 317-318. We agree that risk adjustment – for clinical or social risk factors – is important in the context of performance measurement; however, the clinical measures included in our analysis that medical groups reported to MNCM followed the NCQA-HEDIS measure specifications, which do not call for risk adjustment. Additionally, the data we had for our analyses was aggregated and reported at the medical group level and as such, we could not carry out a person-level risk adjustment model to examine whether the results would be sensitive to adjustment.

The focus of this paper was not to consider how to improve the methodologies used to measure performance. Instead, our aim was to use publicly available data to explore the real-life implications of how different approaches to defining “high performance” affect the classification of medical groups. As such, any limitations pertaining to risk adjustment is a limitation of the data available for this analysis. We have now added and expanded upon this point in a discussion of limitations of the data on Pgs. 14-15, Lines 316-328.
Reviewer #2

1. While the study seems sound, I wondered whether a probabilistic approach (thus the use of statistics) would be needed to explore the role of chance. Differences between the results of various categorization approaches are partly related to 'chance', a concept that covers measurement error, estimation error, and random fluctuations in reality. Thus, I would recommend to involve a statistical advisor in the project.

Thank you for this comment, which we discussed with our two statistical advisors. The issue of error in the measurement of performance measures is an important one. Because MNCM follows NCQA-HEDIS and CAHPS specifications in the measurement of medical group performance, they adhere to minimum denominator requirements for reporting to ensure reliability of results (i.e., more signal, less error or noise in the estimate). For example, NCQA imposes a minimum of n=30 cases to report on a clinical measure performance. Beyond this, though, the publicly available data on the website contained no measure of variability. As an example, the image at the end of the attached response letter shows a sample of the data available to the user: the clinic or group, and the proportion of patients meeting the performance criterion (in this case, the diabetes composite measure). These are simply rank-ordered, and there is no information available or attempt made to distinguish differences in score due to random error.

By delving deep into the documentation on the MNCM website it is possible to find measures of variability for each of the performance measures for at least the year 2015, except for cost, which has no variance reported. Even if we excluded cost from our assessment of provider performance, it is beyond the means of most users of publicly reported performance data to calculate the variance as measures are aggregated into domains, and then again as domains get aggregated into overall performance. Rather, what most users of this information have available to them is what is shown on the MNCM website, namely a rank order of medical groups with some classification as to which are the "top performers" (which in the case of MNCM is the top 15 groups). Our study shows how sensitive an overall assessment of performance is to the domains used to make this assessment and the method used to classify a group as a "top performer" or "high performer".
2. The introduction (as well as discussion) is heavily orientated on the United States and market-type healthcare systems. The paper would improve if a broader international perspective would be taken. Methodological questions regarding performance are also relevant in healthcare systems, which do not have such strong market-type of competition (e.g. the National Health Service in the UK).

Thank you for calling our attention to this important point, particularly since as the reviewer notes, the implications of our findings regarding how providers are designated high-performers extends beyond just the U.S. context and is informative for other countries seeking to advance and improve their healthcare performance assessment approaches. While it is true there isn’t market-type competition in other countries among providers, increasingly other countries are actively measuring provider performance and publicly displaying the provider performance results to help guide consumer choice of provider. We have revised the Abstract (Pg. 2, Line 47 and 67); revised and included additional context regarding performance assessment in non-US healthcare systems in the Background section (Pg. 4; 1st paragraph, Lines 73-83) and made edits accordingly to the Conclusion (Pg. 16, Line 366)

3. Provider performance is a very broad, multi-domain concept and any set of measures has limitations. Particularly the measurement of quality (p.6, 1.125) is fairly limited and focused on population management rather than individual medical care.

We agree with the reviewer that there are important limitations of existing measures of performance, in particular, that measurement schemes today cover only a small swath of all care delivered (although what is measured affects a substantial portion of the patient population). Even the IOM domains available for assessment within the MNCM dataset, and their corresponding measures, represent only a fraction of care provided to patients. We note this limitation in in a discussion of limitations of the data on Page 14, Lines 309-315.
4. Data-analysis (p.8-9): it would be relevant to report if the analyses were pre-specified or iteratively developed during the study. Also, were many more analyses done and not presented?

We pre-specified the conceptual methods for our analyses based on our objective of understanding how different classifications of performance affected the designation of medical groups as high-performing, and we have now added text to clarify this on Page 9 under the Methods subsection “Analyses”. Specifically, we pre-specified: 1) the specific domains we used to define performance (i.e.; quality, cost, access, patient experience) based on the IOM conceptual framework and our prior work (Page 6, Lines 135-140); 2) the use of an “all-or-none” approach to defining high-performance (Page 7, Lines 147-149); and 3) the comparison of relative value versus absolute value thresholds as two approaches to assessing performance based on their common use in performance reporting (noted on Page 9, Lines 201-206). To specify and test different relative and absolute value thresholds we conducted numerous sensitivity analyses to ensure meaningful and comprehensible results. This is described in the Analyses subsection of the Methods section, where we describe both the relative value and absolute value approaches on Pgs. 10-11, Lines 207-240. They are also explicitly presented in Table 1, though we tested more threshold values within the ranges presented. The text in lines 225-227 indicates where we used trial-and-error to find the thresholds for the all-or-none approach that produced meaningful results (“From this initial threshold, we adjusted within each domain to avoid situations where almost all or no groups were high performing on any individual measure, resulting in the following absolute value thresholds”)

5. Results (p.10): It would be relevant to present more information on the provider groups and the database of the performance measures. Currently, it is unclear how many data per provider were available for the performance measurement.

We provide information regarding provider groups and the database of the performance measures in the Methods section. First, on Page 7, Lines 150-158, in the second paragraph of the subsection “Measures”, we note that we selected only those measures within the 4 performance domains that are reported by the most medical groups, to maximize the number of groups that could be included in our study sample, and to ensure that we were including “like” medical groups. We also note that the included measures address care that is commonly provided by both generalist and specialist providers, and which are the focus of measurement and quality improvement efforts of public and private payers and medical groups. We note that we excluded those measures that only some groups reported, such as specialty measures like C-section rates. We also now include an Appendix 1 listing all available measures in the MNCh database and highlighting the ones we selected. Second, on Page 8, Lines 175-180, we note that of the 240 provider groups in the database, 78 groups reported data on all four quality measures and of these 67 also reported data on cost, and of these 58 groups also reported all access and patient
experience measures. Thus, our final sample includes 58 medical groups reporting all selected measures across all 4 performance domains.

6. Discussion (p.12). It would relevant to discuss the issue of meaningful aggregation level. Even a single hospital can be highly heterogeneous in terms of performance (some departments may perform much better than others).

We agree that the level of aggregation is important to address in the context of performance reporting. All data presented in our analyses are presented at the medical group level rather than at the individual clinic or provider level. We chose the medical group as the level of analysis as some measures were only available at the group level (which relates to ensuring the reliability of performance results and provider cost burden associated with collecting, such as the patient experience survey). For measures captured only at the medical group level, we are unable to disaggregate that data to assess performance at the clinic or individual physician level. We now include a description of this in the manuscript in the Methods section on Page 8, Lines 171-173. We caution that measurement at the individual physician level routinely is very noisy and not reliable. As noted in other parts of the manuscript and response letter, we were limited by the publicly reported data available to us on the MNCM website, and thus any limitations related to aggregation level reflect limitations of the data available to stakeholders for decision-making.

Other Revisions

To emphasize that the aim of the current paper was to use publicly-available data to understand the real-life implications of different approaches to classifying performance, we have added more information regarding the Minnesota Community Measurement (MNCM) dataset we used, which is currently used by consumers, payers, and policymakers to make important decisions about providers on the basis of performance (Methods, Page 6, Lines 125-133). We have also added a sentence to the first paragraph of the Discussion section (Page 14, Lines 298-300) to reiterate the aim of the study. We have also made revisions to the last two paragraphs in the Background section (Pgs. 5-6) to emphasize the aim and rationale of the study. Finally, we have made minor edits throughout to increase comprehensibility in the context of the revisions we have made.
Thank you again for giving us the opportunity to revise our manuscript. We hope the changes we have made improve the quality of the manuscript and look forward to hearing from you regarding the revised manuscript.

Sincerely,

Sangeeta C. Ahluwalia, PhD
1776 Main Street
Santa Monica, CA 90407
(Work): 310-393-0411 x7546
(Cell): 510-673-7531
(Fax): 310-260-8151
sahluwal@rand.org