Conference Proceedings

RAND Technical Expert Panel Medicare Advantage (MA) and Part D Contract Star Ratings, November 19, 2020 Meeting

MEETING SUMMARY

Cheryl L. Damberg and Maria DeYoreo

For more information on this publication, visit www.rand.org/t/CFA1147-2

Published by the RAND Corporation, Santa Monica, Calif.

RAND Technical Expert Panel
Medicare Advantage (MA) and Part D Contract Star Ratings

November 19, 2020

MEETING ATTENDEES

Technical Expert Panel Members

NAME	POSITION
Arlene Ash, Ph.D.	Professor and Division Chief for Biostatistics and Health Services Research in the Department of Quantitative Health Services, University of Massachusetts Medical School
Liza Assatourians, J.D.	Vice President of Federal Programs, America's Health Insurance Plans (AHIP)
Kim Caldwell, R.Ph., B.S.	Principal, Texas Star Healthcare Consulting
Darrell Gaskin, Ph.D.	Professor and Director of the Center for Health Disparities Solutions, Johns Hopkins University
Emma Hoo, B.A.	Director, Pacific Business Group on Health
Deborah Paone, Dr.PH., M.H.S.A.	Performance Evaluation Lead for Quality Measurement, Social Determinants of Health, and Care Innovation, Special Needs Plans (SNP) Alliance
Ninez Ponce, M.M.P., Ph.D.	Professor, University of California Los Angeles Fielding School of Public Health's Department of Health Policy and Management, Director of the Center for Global and Immigrant Health
Patrick Romano, M.D., M.P.H.	Professor of Medicine and Pediatrics, University of California Davis School of Medicine
Allyson Schwartz, M.S.S.	President and CEO, Better Medicare Alliance
Jane Sung, J.D.	Senior Strategic Policy Advisor, American Association of Retired Persons
Dolores Yanagihara, M.P.H.	Vice President of Analytics & Performance Information, Integrated Healthcare Association (IHA)

RAND Staff

Cheryl Damberg, Project Director
Maria DeYoreo, Project Codirector
Marc Elliott, Principal Senior Statistician
Melony Sorbero, Senior Policy Researcher
Anagha Tolpadi, Statistical Analyst
Jessica Phillips, Project Manager

MEETING SUMMARY

Welcome and Introductions

- The RAND Corporation meeting facilitator and project director, Cheryl Damberg, began the meeting by welcoming attendees and reviewing housekeeping items.
- Dr. Damberg reviewed the meeting agenda and goals. The RAND team indicated that it would present an overview of different approaches to (1) increase awareness of disparities in performance and (2) strengthen incentives to reduce disparities within the MA and Part D Star Ratings.
    - Specifically, RAND researchers have been exploring two approaches: (1) reporting measures of performance by beneficiary subgroups (i.e., stratified reporting) and (2) assessing performance among those with social risk factors.

Stratified Reporting of Performance

- RAND team member Anagha Tolpadi presented slides about the feasibility of stratified reporting using three Healthcare Effectiveness Data and Information Set (HEDIS) measures reported by four strata.
- A Technical Expert Panel (TEP) member asked for additional information about beneficiaries included in the "other" stratum classification.
    - The RAND team responded that HEDIS defines "other" as a member who has ESRD-only status or is assigned "9—none of the above." The number of beneficiaries that fell in the "other" category was very small. Therefore, RAND researchers excluded this category from analyses.
- A TEP member asked what proportion of beneficiaries fell into the "missing/unknown" stratum?
    - The RAND team responded that, like the "other" category, a very small percentages of beneficiaries fell into this category.
- A TEP member asked for additional context on the number of beneficiaries in each stratum in relation to the denominator minimum requirement of 30 beneficiaries.
    - The RAND team responded that the National Committee for Quality Assurance (NCQA) uses a sample of beneficiaries drawn from the eligible population when reporting performance using the hybrid method and uses all eligible beneficiaries when reporting using the administrative method.
    - The mean and median number of beneficiaries within a contract overall (i.e., regardless of stratum) for the *Breast Cancer Screening* measure, which is an administrative-only measure (i.e., contracts must provide data using the administrative method), was 9,502 and 2,645. The mean and median number of beneficiaries within a contract overall (i.e., regardless of stratum) across the other two measures, which are hybrid measures (i.e., contracts can choose to provide data using either the administrative or hybrid methods), were 751 and 411 (*Colorectal Cancer Screening*) and 1,147 and 411 (*Comprehensive Diabetes Care—Eye Exam*).
    - RAND researchers also noted that NCQA has established a minimum denominator requirement of 30 for publicly reporting performance on clinical measures; CMS follows this criterion for use of the measures within Star Ratings.

- A TEP member suggested making enhancements to the selection criteria used for sampling for hybrid measures. Increased diversity among samples could strengthen the opportunity for stratified reporting among a greater number of contracts. Instead of randomly selecting the sample, CMS could ask that plans randomly select within strata to ensure representation of subgroups.
- A TEP member asked if the *Comprehensive Diabetes Care—Eye Exam* is a composite measure.
    - The *Comprehensive Diabetes Care—Eye Exam* measure is an individual measure that is part of a comprehensive set of diabetes measures. RAND researchers focused their analyses on measures NCQA recommended for stratified reporting. CMS has beneficiary-level data for all measures, so they could expand stratified reporting to more measures. RAND researchers focused their analyses on the three measures presented today to better understand feasibility before expanding to other measures.
    - The TEP member further explained that this diabetes measure is a process measure, and one would expect outcome measures to be more sensitive to disparities. This member suggested that further efforts to evaluate the feasibility and utility of stratified reporting focus on outcome measures.

Discussion Questions About Stratified Reporting

- The RAND team posed the following questions to the TEP:
    - *Given the findings of our analyses, should CMS report contract-level stratified performance results (e.g., confidential to plans, display page, or on Medicare Plan Finder)?*
    - *Should CMS pursue stratified reporting for additional HEDIS measures or non-HEDIS measures?*
    - *Should CMS require a reliability of 0.7 or higher for stratified reporting, or are there scenarios where a lower reliability threshold is acceptable?*
    - *Are there other analyses RAND should perform to inform decisionmaking related to stratified reporting?*
- A TEP member commented that CMS could do more stratified reporting if it looked at the combination of LIS/DE (i.e., those who are dual eligible for Medicare and Medicaid [DE] and/or receive the Part D low-income subsidy [LIS]) or disabled versus neither. RAND researchers agreed that this is one possibility for defining strata.
- A TEP member expressed support for stratified reporting but cautioned that more work needs to be completed to ensure that reporting is as accurate (i.e., with reliable results) and robust as possible. Additional work is also needed to understand what drives performance differentials. The TEP member recommended making stratified reporting confidential to contracts or contract cohorts before making it publicly available.
- There was consensus among the TEP members that stratified reporting is not yet ready to be public-facing but should be shared with contracts.
    - A TEP member noted past work has found confidential reporting does lead to increased buy-in from contracts and is helpful for working out "the bugs."
- A TEP member commented that stratified reporting has value for contracts and is a good a teaching tool but needs to be confidential.

- A TEP member expressed concern about contracts for which there are not enough data to report stratified performance and how that would be interpreted by a lay audience. Public reporting would require significant education efforts with beneficiaries focused on how to interpret contract scores for those that have enough data for reporting versus those that do not, and what it means to not have data displayed. In the interim, stratified reporting should be confidential with contracts.
    - RAND researchers noted that cognitive testing of stratified reporting by race/ethnicity in another context found that when there is missing data, lay people infer the contract does not provide care to many beneficiaries belonging to that race/ethnicity category, which was mainly true. Some participants took the inference further, assuming that contracts with few beneficiaries of a given race/ethnicity were unreceptive to such beneficiaries or performed poorly with them.
- A TEP member commented that if CMS wants to pursue stratified reporting through tools such as the display page or Medicare Plan Finder, the sampling scheme for measures needs to ensure adequate sample sizes across strata (e.g., over-sampling of lower-prevalence strata).
- Another TEP member noted that it would be helpful for NCQA to provide additional guidance to contracts on sampling to ensure different types of beneficiaries are included. NCQA should revisit the literature to identify which risk factors should be targeted.
- A TEP member commented that stratified reporting could be most important for outcome measures. Social risk factors interfere with outcomes. Thus far we have not reviewed results about differences in performance between strata.
    - RAND researchers noted that all three measures included in today's stratified reporting presentation were included in performance results reviewed during the previous TEP meeting on November 12th. The disparities varied in magnitude across contracts.
- A TEP member commented that the minimum requirement of 30 beneficiaries seems low and would result in noisy measurement. In most contracts there is a majority group of beneficiaries, and one could pool all other strata to see if there are differences compared to the majority to get a more powerful signal. If numbers are too small to report by individual stratum, report back the small numbers problem. That being said, sometimes knowing a contract has "a few people like me" is an important signal to a beneficiary.
    - RAND researchers commented that the denominator minimum of 30 is a threshold set by NCQA many years ago for public reporting, to guard against reporting especially noisy estimates. Contracts can have far more than 30 beneficiaries overall and by individual strata.
- A TEP member shared that she works with smaller populations and hears concerns from advocates about suppressing data; therefore, descriptive information should be provided when possible.
    - RAND researchers noted that there is also evidence that when unreliable information is presented it may be interpreted as meaningful even when it is heavily influenced by chance, so there is a trade-off.
- A TEP member stated that moving in the direction of stratified reporting is important and getting to a place where performance can be publicly reported is a

good goal. As an initial step, sharing stratified results with contracts, confidentially, would be impactful. It is important to hold contacts accountable and we do not want to make excuses, but it is hard for contracts to make changes when working with really small numbers.

- A TEP member asked what confidential reporting to contracts would look like. Would reporting include performance on an individual contract basis or include information about all contracts?
    - RAND researchers responded that they were interested in hearing suggestions from the TEP.
- A TEP member recommended complementing stratified reporting with a more holistic approach that characterized the overall association of contract characteristics, such as the association of the proportion of at-risk beneficiaries with performance and disparities.
- Another TEP member agreed that the size of the population within a contract is important. Other important contract characteristics include the size and geographic spread of a contract. A large contract, spanning many states, could be supporting the right initiatives at the contract level; however, providers in different geographic areas are drastically different.
    - A TEP member commented that without knowing the region or size of the contract, the information is meaningless and not actionable.
    - Multiple TEP members shared an interest in learning more about geographic and practice site variability within larger contracts.
        - POST-MEETING NOTE: In previous TEP meetings, the RAND team has shared information about geographic variability.
- A TEP member recommended focusing on measures that address services delivered in the primary care setting. Screening measures often require a second site of care that are impacted by factors such as transportation. Primary care might be more consistent.
- RAND researchers provided additional context about using a reliability cutoff of 0.7. HEDIS measures for contracts that have low enrollment (between 500–1,000) must meet the reliability requirement of 0.7 for inclusion in the Star Ratings. In the past, contracts with fewer than 1,000 enrollees were not scored or included in Star Ratings, but that rule was adapted in order to include more contracts in the Star Ratings while addressing potential concerns about noisy estimates. No such threshold is used for contracts with enrollment of 1,000 or more, which is most contracts.
- A TEP member commented that the reliability criterion should not drop to less than 0.6 for stratified reporting.
- A TEP member expressed support for keeping the reliability cutoff at 0.7.
- A TEP member expressed support for the highest reliability cutoff so people can be confident about reported differences in performance.
- There was broad consensus among the TEP for reliabilities of 0.7 or higher.
- A TEP member expressed dislike of set cutoffs and prefers to make decisions based on the distribution.
    - RAND researchers responded that sample size minimums try to solve a similar problem. The concern is that when information is made public, users assume it is meaningful and reliable or it would not be made public.

- RAND researchers shared three possible social risk factor (SRF) groupings with the TEP.
    - A TEP member recommended collapsing SRF groups only when needed. To the extent reporting is feasible by individual stratum, it should be done.
    - A TEP member noted that the presented groupings have different purposes. Grouping 1 (using the four separate strata) is the best option; however, Grouping 3 (comparing LIS/DE or disabled versus neither SRF) would provide larger numbers.
    - A TEP member agreed: Grouping 1 is preferred, Grouping 3 makes sense if needed.
    - There was less initial support for Grouping 2 (comparing each SRF one at a time—i.e., separately comparing LIS/DE enrollees versus non-LIS/DE enrollees and disabled enrollees versus non-disabled enrollees).
    - It was noted that Grouping 3 makes the strongest assumptions (assumes that different social risk factors have equal effects), Grouping 1 makes the weakest assumptions (does not assume that that social risk factors have similar or additive effects), with Grouping 2 in between.
        - It was also noted that Grouping 1 requires the largest sample sizes and Grouping 3 the lowest.
        - Several TEP members concurred that the choice of grouping should be informed empirically—e.g., Grouping 1 would be supported if social risk factors had different and non-additive effects and if sample sizes were large.
- A TEP member suggested creating contract-level cohorts for comparison, such as a two-way analysis of high/low total enrollment and person-specific groupings (disabled/not).
    - RAND researchers clarified that it is grouping beneficiaries with characteristics within contracts, to look at within-contract variation. If you look at overall performance with differences in population mixes, this would then capture both the between and within-contract variation. Our goal was to focus only on the within-contract variation.

Social Risk Factor Performance Index

- RAND team member Melony Sorbero presented slides on work to construct a Social Risk Factor Performance Index (SRFPI). The goal of the SRFPI would be to incentivize contracts to perform well among socially at-risk beneficiaries.
- The RAND team clarified that the SRFPI would be an addition to the Star Ratings program, not a replacement of the Categorical Adjustment Index (CAI). RAND researchers initially explored applying the SRFPI as a reward, but the SRFPI could also be treated like a measure. The RAND team is continuing to explore different options for construction and use of the index.

Discussion Questions About the Social Risk Factor Performance Index

- The RAND team posed the following questions to the TEP:
    - *Could inclusion of an SRFPI or similar index/measure strengthen efforts to reduce disparities?*
    - *How might the SRFPI be best used?*

        - *To report information on the SRFPI measure to contracts?*
        - *To reward contracts (and what size reward is appropriate)?*
        - *As a measure in Star Ratings?*
    - *Are there specific analyses RAND should perform as it continues to advance the development of an SRFPI?*
- A TEP member commented that ideally contracts that have a large proportion of beneficiaries who are LIS/DE and disabled (relative to the total number of beneficiaries they serve) would benefit most from the SRFPI, but in the examples provided it looks like already high-performing contracts would receive the reward, even if they do not have a large proportion of LIS/DE and disabled enrollees.
    - RAND researchers confirmed no contracts that received at least 3.5 stars on the index received less than 3 stars in the Star Ratings.
- A TEP member expressed a similar concern that contracts that serve a large proportion of beneficiaries who are LIS/DE and/or disabled are disadvantaged in the main Star Ratings, and the index currently does not help much.
    - RAND researchers responded that earlier simulations prorated the size of the reward by contracts' percentage of LIS/DE and/or disabled beneficiaries, which did push the reward towards contracts with a higher percentage of LIS/DE and/or disabled beneficiaries. However, a challenge in applying the index is that large contracts with a modest percentage of LIS/DE and/or disabled enrollees that serve a large number of LIS/DE and/or disabled enrollees did not receive a meaningful reward. So, there is a tradeoff between rewarding contracts serving a large number of LIS/DE and/or disabled beneficiaries or contracts that serve a large percentage of LIS/DE and/or disabled beneficiaries (relative to the total number of beneficiaries the contract serves).
    - The TEP member replied that dollars for disparities reduction should not be spent on contracts that serve low proportions of high-risk (e.g., LIS/DE or disabled) beneficiaries, even if they perform well.
    - Another TEP member strongly agreed with this point; the TEP member stated that dollars should not go to contracts that are compositionally different than those that serve at-risk beneficiaries. Additionally, this TEP member expressed uncertainty about whether measurement should use numbers or percentages of LIS/DE or disabled beneficiaries in a contract. It would be helpful to review a prorated version of applying the SRFPI reward or other variations that might give greater emphasis to contracts serving more at-risk beneficiaries.
        - POST-MEETING NOTE: In the prorated version, a contract that is 100 percent LIS/DE and/or disabled and met the performance criterion for the SRFPI would receive the full reward. Other contracts that met the performance criterion would receive the reward (e.g., 0.5 stars) multiplied by its percentage of LIS/DE and/or disabled beneficiaries. Under this example, a contract that met the criterion and was 50 percent LIS/DE and/or disabled would receive a reward of 0.25.
- A TEP member commented that the goal of the SRFPI is right but in its current state does not move the needle where it needs to move. The SRFPI needs further development.