MEETING SUMMARY

Cheryl L. Damberg and Maria DeYoreo
# RAND Technical Expert Panel

## Medicare Advantage (MA) and Part D Contract Star Ratings

**November 12, 2020**

### MEETING ATTENDEES

#### Technical Expert Panel Members

<table>
<thead>
<tr>
<th>NAME</th>
<th>POSITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arlene Ash, Ph.D.</td>
<td>Professor and Division Chief for Biostatistics and Health Services Research in the Department of Quantitative Health Services, University of Massachusetts Medical School</td>
</tr>
<tr>
<td>Liza Assatourians, J.D.</td>
<td>Vice President of Federal Programs, America’s Health Insurance Plans (AHIP)</td>
</tr>
<tr>
<td>Anne Burns, R.Ph.</td>
<td>Vice President, Professional Affairs at the American Pharmacists Association (APhA)</td>
</tr>
<tr>
<td>Kim Caldwell, R.Ph., B.S.</td>
<td>Principal, Texas Star Healthcare Consulting</td>
</tr>
<tr>
<td>Darrell Gaskin, Ph.D.</td>
<td>Professor and Director of the Center for Health Disparities Solutions, Johns Hopkins University</td>
</tr>
<tr>
<td>Emma Hoo, B.A.</td>
<td>Director, Pacific Business Group on Health</td>
</tr>
<tr>
<td>Eve Kerr, M.D., M.P.H.</td>
<td>Louis Newburgh Research Professor of Internal Medicine, University of Michigan Medical School, Director of the Ann Arbor VA Center for Clinical Management Research, and Director of the Michigan Program on Value Enhancement</td>
</tr>
<tr>
<td>Deborah Paone, Dr.PH., M.H.S.A.</td>
<td>Performance Evaluation Lead for Quality Measurement, Social Determinants of Health, and Care Innovation, Special Needs Plans (SNP) Alliance</td>
</tr>
<tr>
<td>Ninez Ponce, M.M.P., Ph.D.</td>
<td>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</td>
</tr>
<tr>
<td>Patrick Romano, M.D., M.P.H.</td>
<td>Professor of Medicine and Pediatrics, University of California Davis School of Medicine</td>
</tr>
<tr>
<td>Allyson Schwartz, M.S.S.</td>
<td>President and CEO, Better Medicare Alliance</td>
</tr>
<tr>
<td>Jane Sung, J.D.</td>
<td>Senior Strategic Policy Advisor, American Association of Retired Persons</td>
</tr>
<tr>
<td>Dolores Yanagihara, M.P.H.</td>
<td>Vice President of Analytics &amp; Performance Information, Integrated Healthcare Association (IHA)</td>
</tr>
</tbody>
</table>
**RAND Staff**
Cheryl Damberg, Project Director
Maria DeYoreo, Project Codirector
Marc Elliott, Principal Senior Statistician
Carolyn Rutter, Senior Statistician
Melony Sorbero, Senior Policy Researcher
Marika Booth, Statistical Analyst
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 topics covered during this meeting were: (1) providing the justification for addressing disparities in the Medicare Advantage (MA) and Part D contract Star Ratings program, (2) understanding the role of social risk factors in MA and Part D contract performance, and (3) describing approaches for ensuring accurate performance measurement.

Addressing Disparities in the MA and Part D Contract Star Ratings Program
- Marc Elliott, RAND Principal Senior Statistician, presented background information about what social risk factors are, reasons for disparities in care, and work completed by various organizations and researchers to examine the role of social risk factors in performance measurement.
- A Technical Expert Panel (TEP) member commented that there is a lack of consensus about adjustment for social risk factors. Some believe adjustment masks disparities, whereas others believe that failure to adjust masks within-provider (i.e., MA and Prescription Drug Plan [PDP] contract) disparities and that adjustment is necessary for fair comparisons of performance.
  o The TEP member also commented that when you present stratified results by themselves, you lose a sense of the prevalence of the social risk factor across contracts, (for example, whether a contract has 2 percent disabled beneficiaries or 60 percent disabled beneficiaries). It is important when stratifying performance to distinguish between contracts that care for a large versus small proportion of at-risk populations.
- A TEP member suggested including two additional groups of beneficiaries in adjustment: beneficiaries who are “near dual eligible” and beneficiaries who are low income but not eligible for Medicaid. These populations share many of the same social risk factors discussed today.
  o RAND responded that beneficiaries who fall into the low socioeconomic status categories are of interest. The team is seeking the TEP’s suggestions for ways to improve measurement and other data sources to use to measure social risk factors.
- A TEP member noted functional limitations compound poor health outcomes and should be considered for inclusion.
  o RAND responded that the National Academies of Science, Engineering, and Medicine (NASEM) did not include disability as a social risk factor in its work. However, the U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation (ASPE) was interested in disability as a social risk factor and explored the relationship between disability and performance across the various CMS value-based reporting systems.

---

and payment programs as summarized in ASPE’s 2016 report to Congress.\(^2\) Disability has both medical and physical aspects.

- RAND noted that ASPE’s 2020 report to Congress explored the role of functional impairment using survey-based measures. Work is underway to develop claims data–based algorithms that assess functional impairment.

- A TEP member suggested including additional factors in adjustment: frailty, behavioral health, mental health, adverse childhood experiences, and long-term needs. All of these factors interact with health outcomes. The TEP member acknowledged that capturing information about these types of risk factors continues to be a challenge.

### Understanding the Role of Social Risk Factors in MA and Part D Contract Performance

- Melony Sorbero, RAND Senior Policy Researcher, presented information about performance for MA and Part D beneficiaries who are (1) dual eligible for Medicare and Medicaid (DE) and/or receive the Part D low-income subsidy (LIS) and (2) qualify for disability. Dr. Sorbero summarized work completed under CMS’s Part C and D Star Ratings contract with RAND on identifying and addressing LIS/DE and disability disparities in performance measurement.

- A TEP member commented that they would like to see something other than a straight-line regression put through these data (slides 30 and 31 of the presentation) when so much of the data is on one side. Perhaps RAND could consider a locally weighted scatterplot smoothing or something else to capture possible nonlinear relationships.

- A TEP member noted that because Star Ratings are shown in coarse (0.5 star) increments, the sorting into those 0.5 star increments may be more important than the continuous regression relationship (slides 30 and 31).

- The TEP member asked for clarification about whether Medicaid and LIS have similar eligibility thresholds.
  - RAND responded that everyone in the mainland United States who is eligible for Medicaid is eligible for the LIS subsidy (note that Puerto Rico residents are not eligible for LIS). There are individuals who are not eligible for Medicaid but receive Supplemental Security Income and are deemed eligible for LIS. In general, there is a higher income level threshold for LIS than for Medicaid, which helps even out eligibility across states and helps the near poor who are not eligible for Medicaid qualify for benefits.

- A TEP member noted the importance of acknowledging individual choice as a factor in determining whether recommended care is received (e.g., colonoscopy). This member suggested it would be interesting to track patients over time in multiple contracts to see if the individual repeatedly does not receive the right care; this would indicate that it is the choice of the individual rather than a contract issue. The

TEP member stated that the discussion was not using terms that acknowledge individual choice.

- A TEP member asked about the spread shown for each measure on slide 39, because for some measures the spread is narrow while other measures have huge variation. What does RAND make of the variation in the spread of these measures?
  - RAND provided an explanation using the Breast Cancer Screening and Flu Vaccine measures as examples, with possible interpretations for the observed spread.
    - In the case of Flu Vaccine, the median contract immunized its LIS/DE beneficiaries at a rate that is approximately four percentage points lower than for its beneficiaries who are not LIS/DE. This gap in performance is remarkably similar across contracts. Less than five percent of contracts have a gap that is larger than seven percentage points. One possible interpretation is that it is more difficult to achieve a higher rate of immunization with LIS/DE beneficiaries than other types of beneficiaries and that no contract has solved that problem.
    - The rate at which LIS/DE beneficiaries complete breast cancer screening is six percentage points lower than for beneficiaries who are not LIS/DE in the median contract. However, more than five percent of contracts do not have an LIS/DE disparity on this measure. One possible interpretation is that it is harder on average to complete breast cancer screenings for LIS/DE beneficiaries but there is variation across contracts in how they have taken this challenge on, demonstrating that some contracts have been able to address this challenge.

- A TEP member commented that where the 90 percent interval bar includes zero, this suggests some contracts can close the gap.
  - RAND noted that the lack of difference could be that some contracts had closed the gap in performance at a reasonable level or could correspond to low-disparity contracts having uniformly poor performance for both LIS/DE and non-LIS/DE beneficiaries.

- A TEP member noted that the visuals included in the presentation display differences but not the absolute levels. It would be helpful to review results by level because the goal is to achieve high quality care for everyone. The National Quality Forum (NQF), which is examining risk adjustment, proposed peer group comparisons. Examining between-group differences highlights structural reasons for differences. Another TEP member concurred that absolute levels matter.

- Multiple TEP members agreed with earlier comments about the importance of reporting by peer groups (i.e., contracts with high and low proportion of LIS/DE and disabled beneficiaries).

- A TEP member asked about reporting contract performance when two contracts merge into one and taking the higher Star Ratings value from the old contract. If the acquiring contract has enhanced data capabilities, it may help improve scores through better data capture and reporting and targeting high risk beneficiaries.
• RAND noted that how CMS handles consolidated contracts changed starting with the 2020 Star Ratings. Now if two contracts merge, they receive a weighted average of the two contracts’ performance when computing the Star Rating.
• For measure-level adjustment, a TEP member emphasized the need for minimum standards for measure testing and guidance to measure developers about how to test for the role of social risk factors. Other TEP members agreed with this comment.

Approaches to Ensuring Accurate Performance Measurement
• Melony Sorbero and Anagha Tolpadi of RAND presented two approaches to ensure accuracy of measurement of performance: (1) direct adjustment through case-mix adjustment and (2) indirect adjustment through the Categorical Adjustment Index (CAI).
• A TEP member asked for additional clarification about the categories used in the CAI.
  o Contracts are grouped into ten categories (deciles) using percent LIS/DE and five categories (quintiles) using percent disabled.
  o The TEP member expressed concerns about the groupings given the wide spread of percentages within some of the categories. For example, category five for disabled runs from 40 percent to 100 percent disabled, and the characteristics of these contracts could differ significantly. They asked whether it is appropriate to group contracts together that fall at the low and high ends of these categories. For example, the highest disability quintile includes a very large spread of percent disabled.
    ▪ RAND responded that this approach, because it does not assume a linear relationship between risk factors and stars, does require a minimum number of contracts in each category. If the gold standard comparison is a case-mix adjustment model with a fixed effect for LIS/DE, more approximation is taking place in the upper bins.
• A TEP member asked if Spanish or limited English proficiency is included as an adjuster?
  o RAND responded that neither Spanish preference nor limited English proficiency is an adjuster for measures in the Star Ratings program or the MA and Prescription Drug Plan Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey. For the hospital and hospice CAHPS surveys, survey language and/or respondent’s home language is used as an adjuster.

Discussion Questions
• Questions posed to the TEP were as follows:
  • Is the CAI beneficial to include in the Star Ratings?
    ▪ The Medicare Payment Advisory Commission (MedPAC) and ASPE have suggested rescinding it.
  • Does the CAI meet its intended goal of improving accuracy of measures of performance for MA and PDP contracts?
  • Are there ways the CAI could be improved?
• A TEP member commented that the CAI is meeting the goal of reducing bias for contracts with a high proportion LIS/DE and disabled beneficiaries. This member asked for clarification about what ASPE’s and MedPAC’s arguments are for removing adjustment. This member stated that adjusting for social risk factors is appropriate for these clinical measures.
- RAND stated that ASPE supports only very limited use of adjustment. During the next TEP meeting, this issue will be explored further. ASPE work for the 2020 Report to Congress was primarily using analysis of measures from the inpatient setting. The opportunity for things like transportation to impact outcomes matters less in the inpatient setting compared with makingfollow-up appointments.

- ASPE’s recommendation coming out of their 2016 report to Congress was that adjustment should be assessed on a measure by measure basis. ASPE’s 2020 Report to Congress came out with a firmer set of recommendations that did not support adjustment for the vast majority of measures. For quality reporting specifically, ASPE supports adjustment for survey-based measures, while for value-based purchasing programs they support adjustment for survey-based measures and cost measures, not for process measures or outcome measures in general.

- This is one aspect of the 2020 ASPE report that differs from the first ASPE report published in 2016, the NASEM report on methods to address social risk factors, and the NQF report.

- RAND also noted that the recommendations of the 2020 ASPE report do not reflect new findings or analyses in the MA setting and may rely on inpatient measures that are less applicable in this setting.

- A TEP member expressed appreciation for the overview provided today. The CAI is an important aspect of the MA Star Ratings program. CMS should consider possible improvements to the CAI but not elimination of the CAI. Other risk factors beyond LIS/DE and disability should be considered for inclusion in the CAI. Members of the TEP flagged other social risk factors that should be considered that impact performance, beyond DE/LIS and disability.

- A TEP member acknowledged more work is needed to address disparities from the ground up, which will take time and resources across the health care system. According to this TEP member, the CAI should not be eliminated until there are long-term solutions in place. Furthermore, ASPE’s second set of recommendations in the 2020 Report to Congress to rescind the CAI is problematic. The ASPE report does not sufficiently explain or justify the recommendations. Not adjusting may mask inequities, a topic that was highlighted at the start of the meeting. Work should be done to improve the CAI, but not eliminate it.

- A TEP member agreed with prior commenters that the CAI has limitations and noted that it was supposed to be an interim solution, and the TEP member continues to view it as an interim solution. The CAI has potential and does recognize the differences that arise from population differences. There are other things that could

---


4 National Academies of Sciences, Engineering, and Medicine, 2016.

be done, so it does not quite meet the goal of improving accuracy in measurement (per the question to the TEP on slide 63). The TEP member suggested a long-term solution of addressing adjustment at the measure-level, which could be done. At a measure level, measure developers should test their measures in low income, diverse, non–English-speaking populations, etc., using characteristics that have been empirically shown to relate to differences in performance independent of health plan or provider actions. If measure developers could have more guidance on the testing and the methods (sampling, how to accommodate these individuals) and specificity around measure specifications (who to include and when), this would address this within the measure specification to recognize these other characteristics. There needs to be more guidance for and consistency across measure developers when testing measures to ensure the inclusion of diverse, low socio-economic status, non–English-speaking populations of beneficiaries. This would be additive, not mutually exclusive; one could do a bit better on the CAI and modeling different variables while improving adjustment at the individual measure level.

- A TEP member noted that medication adherence measures are not true clinical process measures because they depend heavily on patients’ actions and engagement, and social networks that support those patients. In that way, many of the Star Ratings measures are very different from traditional inpatient process measures. The role of patients in these measures may increase the importance of accounting for social risk factors.

- A TEP member commented that the implementation of the CAI seems like a good approach for not penalizing providers or contracts and motivating them to enroll certain types of patients. However, the TEP member suggested using the CAI in combination with incentives to improve quality.