

RAND Medicare Advantage (MA) and Part D  
Contract Star Ratings Technical Expert Panel  
November 7, 2019 Meeting

**MEETING SUMMARY**

Cheryl L. Damberg and Maria DeYoreo



For more information on this publication, visit [www.rand.org/t/CF417](http://www.rand.org/t/CF417)

Published by the RAND Corporation, Santa Monica, Calif.

© Copyright 2020 RAND Corporation

**RAND**® is a registered trademark.

#### Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit [www.rand.org/pubs/permissions](http://www.rand.org/pubs/permissions).

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

[www.rand.org](http://www.rand.org)

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

November 7, 2019

## MEETING ATTENDEES

### Technical Expert Panel Members

NAME	POSITION
Arlene Ash, PhD	Professor and Division Chief for Biostatistics and Health Services Research in the Department of Quantitative Health Services, University of Massachusetts Medical School
Liza Assatourians, JD	Vice President of Federal Programs, America's Health Insurance Plans (AHIP)
Jane Sung, JD	Senior Strategic Policy Advisor, American Association of Retired Persons
Lindsey Copeland, JD	Policy Director, the Medicare Rights Center
Amy Nguyen Howell, MD, MBA	Chief Medical Officer, America's Physician Groups (APG)
Deborah Paone, DrPH, MHSA	Performance Evaluation Lead for Quality Measurement, Social Determinants of Health, and Care Innovation, Special Needs Plans (SNP) Alliance
Ninez Ponce, MMP, PhD	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, MD, MPH	Professor of Medicine and Pediatrics, University of California Davis School of Medicine
Allyson Schwartz, MSS	President and CEO, the Better Medicare Alliance

### **Technical Expert Panel Members unable to attend:**

Anne Burns, RPh, American Pharmacists Association  
 Emma Hoo, Pacific Business Group on Health  
 Eve Kerr, MD, MPH, University of Michigan  
 Elisa Munthali, MPH, National Quality Forum  
 Dolores Yanagihara, MPH, Integrated Healthcare Association  
 Kim Caldwell, BS, Texas Star Healthcare Consulting, LLC

### **RAND Staff**

Cheryl Damberg, PhD (Project Director)  
 Maria DeYoreo, PhD (co-Project Director)  
 Marc Elliott, PhD  
 Justin Timbie, PhD  
 Andy Bogart, MS

Erin Taylor, PhD  
Melony Sorbero, PhD  
Chau Pham, MS

## MEETING SUMMARY

### Welcome and Introductions

- The RAND meeting facilitator and project director, Cheryl Damberg, began the meeting by welcoming attendees and introducing the new Technical Expert Panel (TEP) member, Emma Hoo.
- Cheryl Damberg reviewed the agenda and meeting goals. In this meeting, RAND requested the TEP's input on: 1) consideration of new measures related to end-stage renal disease (ESRD) or chronic kidney disease (CKD) treatment; 2) generic prescribing measures; and 3) reporting at the parent organization level. RAND provided an update on the analyses around geographic reporting conducted since the April 2019 TEP meeting. *A copy of these presentations can be found in the corresponding slide deck.*

### ESRD Measures

CMS is considering adding ESRD or CKD treatment measures to Part C and Part D Star Ratings in the future. The 21<sup>st</sup> Century Cures Act (CURES; P.L. 114-255) removes restrictions on the ability of Medicare beneficiaries with ESRD to enroll in any MA plan starting in 2021. The extent to which this will result in beneficiaries with ESRD enrolling in MA plans is unknown.

The Part C & Part D Star Ratings currently include only one measure that focuses on kidney disease: Diabetes Care – Kidney Disease Monitoring. ESRD is explicitly excluded from measures on statin use among people with cardiovascular disease and diabetes, and medication adherence for diabetes, hypertension, and cholesterol. In addition, ESRD meets the advanced illness portion of the advanced illness and frailty exclusion for breast cancer screening, colorectal cancer screening, diabetes care – eye exam, diabetes care – kidney disease monitoring, diabetes care – blood sugar controlled, and rheumatoid arthritis management.

RAND described the policy goals identified in the recent executive order on Advancing American Kidney Health, which can be used to guide measure development and selection.

RAND identified 75 existing ESRD or CKD measures. Only six of these measures include health plan as the unit of measurement:

1. Chronic Kidney Disease: Monitoring Parathyroid Hormone (PTH);
2. Chronic Kidney Disease - Lipid Profile Monitoring;
3. Chronic Kidney Disease with LDL Greater than or equal to 130 – Use of Lipid Lowering Agent;
4. Influenza Immunization specific to those with ESRD/CKD;
5. Chronic Kidney Disease: Monitoring Calcium; and
6. Chronic Kidney Disease: Monitoring Phosphorus.

Measures of clinical processes of care were most common (49 measures), followed by outcome and intermediate outcome measures (21 measures). A small number of structural (3), patient

experience (1), and cost/resource use measures (1) were identified. Some of the identified measures address care already assessed in Star Ratings but limit the denominator to those with ESRD or CKD.

Questions posed to the TEP were the following:

- ***Should CMS consider adding ESRD or CKD measures to the Star Ratings in the future?***
  - ***Are there certain ESRD or CKD quality of care measures that should be higher priority for potential inclusion in the Star Ratings?***
  - ***Are there issues CMS should consider in adding ESRD or CKD measures given the uncertainty about how many beneficiaries with ESRD will select MA?***
- 
- One TEP member stated that it seems that one of the most high-value things that one could encourage a plan to do is to prevent or slow the decline from CKD into ESRD. Of all the things that one might look at, that area seems to be most important. It would be important to figuring out how to measure that.
  - A TEP member commented that measuring quality of care for patients with ESRD is a very interesting; however, it remains to be seen how many ESRD patients will opt into MA plans. The majority of the measures in this space are micro-process measures that are measured at the ESRD facility level or dialysis facility level, and, as such it is unclear what the health plan locus of control is for influencing performance on these measures. For health plan ratings, one would want to be looking at measures that have more to do with coordination of care and the overall progression of the patient's condition and not these micro-process measures that relate to the day-to-day management of hemodialysis.
  - Another TEP member agreed that measuring ESRD quality of care should be on the horizon and that it aligns with CMS's Comprehensive ESRD Care (CEC) model. The TEP member commented that ESRD will be a foundational model for value-based care. Possible other measurement topics could be Nonalcoholic steatohepatitis (NASH) and liver disease because these are two emerging diseases and condition states that will significantly affect a lot of Americans, especially as one looks at Medicare Advantage. The TEP member agreed that influenza immunization for the ESRD population is something to consider as well as screening for clinical depression, which aligns with the current set of CEC quality measures. The TEP member recommended as a potential measure having an advance care plan. The TEP member also noted a focus on total cost of care. It would be helpful to look at outcome measures and how these measures affect care coordination or vice versa. The TEP member was interested to see which of these measures were identified as having any linkage to care coordination or transitions of care.
    - RAND commented that they would circulate a list of the measures RAND identified after the meeting and would welcome feedback from the panel on any of the existing measures.
  - A TEP member agreed that it is useful to look at what measures already exist and what is already being measured regarding care as renal kidney disease progresses. The TEP member noted that the relationship between ESRD facilities and how they will coordinate care for other conditions for a person who has ESRD is unknown. Sometimes a patient has other conditions that led to ESRD, and these other conditions do not all go away, or the patient might have other care needs (e.g., preventive care, such as flu vaccine, or management of depression). Knowing what is already measured that is not ESRD facility-specific, particularly for people who are on a progression to ESRD, is certainly something that should be examined before adding new measures. The idea is to consider what does

not exist now and whether care coordination measures or transition of care measures exist for any patients, let alone for someone with ESRD.

- A TEP member commented that if plans have until January 2021 to implement, they need to have some guidance on what the expectations are. The TEP member asked if there is any expectation on the part of CMS that there would be some further discussion about this sometime between now and June.
  - CMS replied that this ESRD discussion is to consider measures that may be added several years from now, not in the next two years. For any measures considered, CMS would need to test them at the plan level to determine if plans have enough ESRD patients to generate reliable estimates of performance. This work would need to be performed before CMS proposes any measures through the regulatory process.
- A TEP member commented that it is helpful to know that there will be opportunities to learn about these quality measures, to test them, and to get feedback from the stakeholders.

## **Generic Prescribing Measures**

CMS is considering additional measures to lower spending in the Medicare program. One approach to reduce prescription drug costs is to increase dispensing of low-cost generic medications, instead of more-costly branded drugs. CMS has constructed two measures of generic prescribing. The first measure, the Generic Dispensing Rate (GDR), measures the proportion of all drugs dispensed that are generic. The second, the Generic Substitution Rate (GSR), measures the proportion of all drugs for which generic alternatives exist that are dispensed as generic; this measure assesses the potential for substitution away from higher-cost branded drugs. Generic prescribing rates have historically been very high in Part D: the GDR for 2017 was 83.5% for MA-PDs and 81.9% for PDPs, and the GSR was 91.6% for MA-PDs and 90.7% for PDPs. Although the overall rates of generic prescribing are very high, rates within certain therapeutic classes may present room for improvement.

Questions posed to the TEP were the following:

- ***What are some pros and cons of incorporating generic prescribing measures into the Star Ratings?***
- ***Should CMS consider any additional generic prescribing measures, such as measures focused on specific therapeutic classes?***
- ***What classes might be worth exploring?***
- ***How could biosimilars best be incorporated into the measures?***
- A TEP member commented that the question of biosimilars is very interesting. There is broad interest in figuring out how to better incentivize the use of biosimilars and this might be one way. The TEP member encouraged everyone to look at some of the reasons why biosimilars are not being used anywhere nearly as much as generics. These quality measures have been so good at incentivizing behavior change, so maybe this is one way to do it for biosimilars, but these strategies won't be effective if the problem is a lack of cooperation with the manufacturers.
- A TEP member stated that this is an open question for discussion and that it gets back to what the health plan controls, whether there would be restrictions or conflicts with the formularies that are established with the providers serving the beneficiary, and how much

they can dictate this action. If the primary prescriber justifies the use and the generic or the biologic similar cannot be found, what then happens? Whether to include a generic prescribing measure depends a lot on what the measure is.

- One TEP member commented that in the Part D program, beneficiaries are either in plans with a specific formulary that tightly controls which drugs are prescribed or plans where there is a difference in the beneficiary's out-of-pocket costs between generic and branded drugs. In both situations, there is a pretty strong financial incentive already in place for most of these beneficiaries to choose or to request a generic drug. If that is true, there may be relatively little opportunity to move the needle. In some cases, generics are becoming just as expensive as brand drugs because of how generics get rebranded by companies that have a monopoly in that market. The TEP member wasn't sure how prevalent that is across the whole spectrum, but it can no longer be assumed that the generic is always cheaper than the brand drug. In some cases, they are actually quite similar in price.
- A TEP member commented that this does not feel like a quality measure. The plans and the individual patients have plenty of financial incentives to choose the cheaper drug, and it is not clear that it is higher quality to have the maximum amount of generic usage.
- Another TEP member agreed that this type of measure is more about total cost of care because the whole point of brand vs. generic drug use is around cost control as opposed to quality of care. The TEP member noted that there is a recent evidence-based study released this year in March regarding the overall generic vs. brand prescribing rate for cardiovascular, diabetes, and psychiatric medications. It is unclear, from the patient advocacy perspective, whether this measure is assessing quality of care or total cost of care.
- One TEP member commented that it is important to take into consideration some of the publications that associate higher use of generic prescribing with quality and noted they knew of at least one study that reported an association between use of generic drugs and higher ratings of member patient experience.
  - RAND agreed that these are important points and will go back and look at what is known about generic prescribing regarding its relationship to cost as well as quality.

## **Reporting Star Ratings at the Parent Organization Level**

RAND presented material on the topic of measuring and reporting Star Ratings at the parent organization level. The majority of parent organizations (63%) in 2019 had only a single contract. However, there are a small number of parent organizations that have many contracts (3% of parent organizations have more than 10 contracts). Currently, most Star Ratings measures are measured and reported at the contract level; other levels for reporting performance could be considered, such as the parent organization level.

RAND gave two examples of measures to illustrate what reporting at the parent organization level would mean, discussing methods for calculating parent organization level scores and how this change in methodology would affect the scores received by contracts. In the case of the breast cancer screening measure, single-contract parent organizations would simply receive the contract score as currently calculated, whereas multi-contract parent organizations would receive scores calculated as a weighted average of contract scores, with all contracts in the same parent organization receiving the same parent organization score. For the disenrollment measure, one approach to reporting disenrollment at the parent organization level would involve a change in the measure specification to reflect the percent of members who voluntarily changed contracts, with a

disenrollment counting as a numerator positive event only if the beneficiary switched to a different contract operated by a different parent organization than their current contract. This would mean that contracts in multi-contract parent organizations would see reduced disenrollment rates if members move between contracts in the same parent organization.

Questions posed to the TEP were the following:

- ***Should we assess variation in scores across contracts within the same parent organization for a subset of measures?***
- ***What are the pros and cons of calculating scores at the parent organization level and what should we consider?***
- ***Which measures, if any, warrant parent organization level reporting?***
- ***Are there particular analyses RAND should consider doing to inform CMS's decisions?***
  
- A TEP member commented that it would be interesting to look at this by separating out whether the contract or the parent organization is responsible for the action, and whether it makes a bigger difference using HEDIS measures vs. using other types of measures (e.g., customer service). Also, some of these performance areas are really functions of the parent organization and some of them are more functions of the contract. It would be useful to know where the variation occurs. They said it makes a big difference in enrollment or disenrollment, but it would really be different if you were measuring clinical care. Who is responsible? More information on the source of the variation would be helpful. Examples are really important. It really depends on which actor conducts the care or administrative service.
  - RAND commented that it is not always clear how one would know for any given parent organization where those functions reside. Some plans may choose to centralize the functions while others may not.
- A TEP member agreed that where actions occur does vary. In fact, as more and more provider groups assume risk and assume functions, that is going to keep changing, too. It is really very different within different parent organizations and varies as to how much they have passed along some of these functions and responsibility.
  - RAND stated that, arguably, this is also an empirical question. One could see empirically how much variation is at the contract level, for example, for each of the measures.
- Another TEP member agreed that this is an empirical, statistical question. The consumer will always want to have information that is drilled down to the unit that is closest to them, but there is a reliability-validity tradeoff because it may be that when you drill down to the contract level, the score level reliability is really poor because you have smaller units. On the other hand, if there is true variability across contracts within a parent organization, then it is important to identify and expose that variation. One could empirically explore whether the parent organization or the contract organization matters more. And it is a nested problem because the contract organizations are nested within parent organizations.
  - RAND added that because reporting is generally at the contract level, the measures have been assessed for reliability at the contract level and that reliability is, generally, pretty good for most contracts.
- A TEP member commented that there are about 3,800 plan benefit packages (PBPs) and that is what a Medicare beneficiary looks at when selecting a plan product. The beneficiary is looking at one of those 3,800 PBPs, not hundreds and not at the parent organization level, not at the contract level, but at the PBP level. The TEP member emphasized that the beneficiary is evaluating the PBP to determine if it is for disabled people, for dually-eligible

people, provides special supplemental benefits for the chronically ill, is in a different provider network, and what the monthly premium and copay are. All of this is at the PBP level, not at the contract level, and definitely is not at the parent organization level. It is important to remember what beneficiaries are looking at and if quality measurement is important to them. The contract level is already challenging, and if one goes up to the parent level, the meaning is gone.

- A TEP member added that these PBP level differences are not infinite and fall into different categories. The TEP member added a useful analysis might be to examine beneficiaries who sign up for certain kinds of PBPs but often switch out as soon as they get sick; such information could be useful to the consumer. The TEP member thought that for most purposes, the higher parent organization level of measurement is just a theoretical construct and this TEP member did not see its use. This level of measurement would certainly not be useful for consumers. The TEP member asked to what extent do we think that at the more aggregated parent organization level are there any of the measures in the Star Rating program that are really under the control at the parent organization level? Potentially there is something that could be learned by doing analyses to see whether there is variation and at what level the variation is occurring. At a measure level, one might find that there are parent organizations that do a better job and for others it is pretty scattered, and all the variability occurs lower down at the contract or PBP level.
- Another TEP member agreed on having the measurement at the contract and not at the parent organization because it might dilute some of the data and the whole validity of the Star Ratings measures. The variation between contracts is important to see because a parent organization can have multiple contracts across multiple states. The demographics, the social determinants, the make-up of that contract and the patients are very important to look at, at that level versus just at the overall aggregate parent organization level.
- A TEP member asked if someone switches from one contract or PBP to another, within the larger parent organization overarching the contract or PBP, whether it might be evidence that the structure of the parent organization entity (i.e., contract or PBP) in which they were enrolled did not suit them. The TEP member would be nervous about counting switching within parent organizations as a non-switch if a beneficiary switches to a different contract within the same parent organization.
  - RAND provided an example of switching from a more expensive to a less expensive option where cost is the only factor and if it happened to involve two contracts that were in the same parent organization, that cost-based decision does not count as a disenrollment, but if it happened between two contracts in two different parent organizations, that cost-based decision gets treated as a disenrollment in the scenario described in the slides.
- The TEP member agreed that counting movement within parent organization as a non-switch would feel wrong since this larger structure is fairly invisible to the individual and the TEP member was unsure why this type of switch should count differently.
- A TEP member stated that given that there are so many different measures, it seems like one would have to deal with the practicality issue. Each one might feel different as to what matters at what level; that is where the empirical information could be helpful. Doing mammograms or not doing mammograms is comparable no matter what level one is talking about, and as such, why not measure and report at the most granular level (i.e., PBP level). On the other hand, some of the other measures seem more problematic in terms of getting to too low of a level because you will have much smaller numbers of beneficiaries in the denominator. It is really a compromise as to what the right level is. It seems like picking the one in the middle (i.e., contract level), which is the current level of measurement and reporting, may be fine, unless CMS wants to go through each one and create each one at the

right level. CMS would not really have any way of knowing what level of the organization (PBP, contract, or parent organization) is responsible. It would be difficult to know where the locus of control is. We want the consumer to have the correct information as close to them as possible, but some of these are not handled similarly. Some things are handled at the contract level and some things are handled at the parent organization level, and each parent organization probably does it a little bit differently and they might even do it differently for the contracts let alone for the PBPs. It sounds like making a big change here is not worthwhile.

- A TEP member noted when looking at the Part C and Part D measures and at what is known about how health plans work and how care systems work, there are a few things like a call center, or the foreign language interpreter, or reviewing appeals decisions and the timeliness of appeals, that may be enterprise-wide at the parent organization level, but it is not certain. It is unknown how much of these functions are centralized in every kind of Medicare Advantage plan and this would require more study. Things around receipt of health care, like monitoring physical activity, measuring body mass index (BMI), flu vaccine, breast cancer screening, bladder control, falling, are very dependent on the care systems, the providers, the region, and the practice patterns and it is very hard to think that nationally there is one standard that somehow the health plan enforces throughout its provider network. It is very hard to see that that is realistic.

## **Update on Geographic Reporting Analyses**

RAND presented the results of descriptive analyses of contract-level variation in performance across states. These new analyses were intended to answer the question of whether beneficiaries would be expected to change plans under state-specific measurement of contract performance. For each contract, RAND examined performance within each state relative to all other contracts in the state and used within-state percentile scores to determine whether a contract's state-specific units were high- or low-performing. RAND examined consistency in performance for each contract across states by displaying each contract's range in performance on each of four quality measures. RAND found that, within a contract, the relative performance of individual geographic units varied widely across each of the measures examined. These differences persisted, but were somewhat smaller, when examining average performance on multiple measures simultaneously. Since a contract's relative performance varies across geographic units, one would expect some beneficiaries to switch to higher performing plans under state-specific measurement. RAND discussed several operational and policy challenges associated with geography-specific measurement, such as beneficiary sample sizes, the incorporation of administrative measures, and potential unintended consequences associated with geographic measurement.

- A TEP member stated that certain Star Rating measures like appeals and call centers do make sense at the contract level, but the other more clinical, preventive health-related measures are probably related to the population enrolled, the characteristics, the network, the PBP. The special needs plans, the Medicare-Medicaid Plans (MMPs), deal with states because they are dealing with dually-eligible individuals, those beneficiaries, those eligibilities, what is covered, what is not, and this makes a big difference in how people utilize health care, and so that does speak to having something that is focused state by state.
  - RAND responded that a broader set of quality measures was not explored. The approach was to take a market basket that covered different types of measures, but also

measures that differed in their likely sample size at smaller units of reporting. For example, CAHPS measures have smaller sample sizes than HEDIS measures. RAND was focused on an illustrative set of measures to see if there is a signal that is unique to geography and separate from contract overall performance.

- The TEP member added the findings were not surprising, particularly because as measures are added or pooled, there is less variation, whereas if one looks measure by measure at the cut points and the trends, one can see that everybody performs highly on some measures and others are challenging for everyone.

- A TEP member stated that the idea is that one is trying to produce something for the consumer that would be relevant to their choice, as they are in a specific geographic area and they have a limited set of competitor plans. The TEP member proposed that there could be a measure in which one compares performance using a weighted approach to the performance of the set of plans in which other people residing nearby are enrolled. So realistically, in certain parts of Massachusetts, for example, consumers might have a very specific set of 2 or 3 plans in which they could enroll while others just are not realistic. Consumers are not going to travel across the state to go to the other plans offered in the state. The place consumers go to get care feels useful because they do not care if the plan performs well in California when they live in Delaware. On the other hand, the state might not be the right unit.

-RAND noted that was a good point. In an earlier analysis, one level of geography below the state, the hospital referral region (HRR), was used. At that level, we start to lose sample within those units, within each contract, and really could not discern whether there was meaningful variation at that level between contracts. So, that was a level that was a little too fine-grained.

- The TEP member further commented that maybe that teaches us something. Maybe the finding that one cannot do it at the level that would really matter suggests that maybe one should not try to do it.
- Another TEP member asked if it matters whether it is a small state or a big state. That is, for small New England states, the relevant geography could be the entire state. Clearly for California or for large states like New York, Texas, or Florida, state-level reporting is kind of silly.
- A TEP member stated that one could argue that the HRR is kind of close to the concept of what is being talked about, but that is where it gets difficult to do. It may be that once we cut down to small enough areas, that is where we start running into the problem of sample sizes.

-RAND added that in some bigger states RAND tried to break states into regions like northern and southern California, but we start running into small numbers unless, somehow, we are able to increase sample size. In surveys you can add sample, but it incurs a cost to whomever is paying for the survey. With HEDIS measures you cannot always increase the denominator as there are only so many beneficiaries eligible for the measure. RAND is aware that more regionalized information could be highly beneficial to the consumer; however, there are some constraints in trying to move into this space that have not fully been sorted out.