February 10, 2025

Ing-Jye Cheng
Acting Director, Center for Medicare
Centers for Medicare and Medicaid Services
Department of Health and Human Services

Re: Advance Notice of Methodological Changes for Calendar Year (CY) 2026 for Medicare Advantage (MA) Capitation Rates and Part C and Part D Payment Policies [CMS-2024-0360]

Dear Ms. Cheng:

Thank you for the opportunity to comment on the 2026 Part C and D Advance Notice published by the Centers for Medicare and Medicaid Services (CMS).[1] In the Advance Notice, CMS states that it may begin calibrating its risk adjustment models using MA encounter data rather than Traditional Medicare (TM) claims data. Unfortunately, this change would not achieve what CMS seems to hope it would—and could have the opposite of the desired effect. I have four main points:

- **Contrary to what CMS suggests in the Advance Notice, a coding pattern adjustment would remain necessary if the risk adjustment model were calibrated using MA encounter data:** In the Advance Notice, CMS states that calibrating the risk adjustment model using MA encounter data instead of TM claims data would "remove the need to make the adjustment for coding pattern differences" under section 1853(a)(1)(C)(ii) of the Social Security Act. In fact, such an adjustment would remain necessary.

  The need for a coding pattern adjustment ultimately stems from the fact that risk scores for MA and TM enrollees are calculated using different source data; risk scores for MA enrollees are calculated using diagnoses reported by MA plans, while risk scores for TM enrollees are calculated using diagnoses reported by providers on TM claims. It is well-established that MA plans report more diagnoses for MA enrollees than would be reported on TM claims.[2] Thus, comparing raw risk scores calculated in this way overstates the claims risk of MA enrollees relative to TM enrollees. Consequently, making valid risk-adjusted comparisons between MA plan bids and TM-based benchmarks, as the current MA payment system requires, necessitates a coding pattern adjustment.

[1] The views expressed in this letter are my own and do not necessarily reflect the views of the Brookings Institution or anyone affiliated with the Brookings Institution other than myself.

[2] Medicare Payment Advisory Commission "Medicare Payment Policy," March 2024, https://www.medpac.gov/document/march-2024-report-to-the-congress-medicare-payment-policy/.

Nothing about this rationale for a coding pattern adjustment would change if the risk adjustment model were calibrated using MA encounter data instead of TM claims. As long as risk scores are calculated from data that capture more diagnoses for MA enrollees than comparable TM enrollees (and more diagnoses translate into higher risk scores, as would surely remain the case), it will still be necessary to adjust MA risk scores downward to allow apples-to-apples comparisons between plan bids and TM-based benchmarks.

- **Calibrating the model using MA encounter data could either increase or decrease the appropriate coding pattern adjustment, depending on how the spending of "upcoded" beneficiaries compares to their peers:** To see why, it is instructive to consider two illustrative scenarios for how the spending of "upcoded" beneficiaries (i.e., those who are assigned a diagnosis when enrolled in MA but not in TM) compares to their peers.

  One scenario is that the upcoded beneficiaries have spending similar to others who are not coded as having the condition in TM; this would be the case if plans' coding efforts are identifying beneficiaries who technically satisfy the criteria for a diagnosis but do not, in fact, have meaningful health care needs associated with the condition. In this case, calibrating the model using encounter data would reduce the appropriate coding pattern adjustment. This is because risk scores for people with the relevant condition would fall (since this group would now include more people with relatively low spending), but risk scores for people without the condition would remain stable (since the people shifted out of this group would be similar to those left behind). Thus, being assigned the diagnosis would cause a smaller increment to risk scores, mitigating the extent to which higher coding intensity in MA translates into higher risk scores.

  Another scenario, however, is that the upcoded beneficiaries have spending similar to people who were already coded as having that condition in TM; this could be the case if plans are identifying conditions that have a major effect on their health care utilization but that often go unrecorded in TM due to the weak incentives for complete coding in TM. In this case, calibrating the model using encounter data would tend to *increase* the appropriate coding pattern adjustment. This is because risk scores for people with the relevant condition would remain stable (since the people added to this group would be similar to those who were already there), while risk scores for people without the condition would decline (since this group would have lost some people with relatively high spending). Thus, having the diagnosis would cause a larger increment to risk scores, magnifying the degree to which higher coding intensity in MA translates into higher risk scores.

  Reality likely lies in between these two extremes. Upcoded beneficiaries likely have lower average spending than beneficiaries assigned the relevant diagnoses in TM; some diagnoses reported by MA plans appear to be invalid,[3] and even where MA plans do identify additional valid diagnoses, it is plausible that these diagnoses often reflect less

[3] United States Department of Health and Human Services, "Fiscal Year 2024 Agency Financial Report," November 14, 2024, https://www.hhs.gov/sites/default/files/fy-2024-hhs-agency-financial-report.pdf.

severe versions of those conditions. However, upcoded beneficiaries likely also have higher average spending than beneficiaries who are not assigned the relevant diagnoses in TM; the overall difference in coding intensity between MA and TM appears to be much larger than can be accounted for by invalid diagnoses,[4] which suggests that many of the additional diagnoses are "real" diagnoses likely to be predictive of higher spending. Thus, without additional empirical analysis, it is not clear whether changing the calibration dataset would increase or decrease the appropriate coding pattern adjustment on net. The appendix presents a mathematical model that makes this point more formally.

- **CMS would bear a heavier evidentiary burden if it wished to apply a coding pattern adjustment after calibrating the model using MA encounter data, and it could face challenges to its authority to apply any such adjustment:** Section 1853(a)(1)(C)(ii)(IV) of the Act states that the statutory requirement to apply a coding pattern adjustment (and the associated statutory minimum adjustment) apply only "until the Secretary implements risk adjustment using Medicare Advantage diagnostic, cost, and use data." The discussion in the Advance Notice indicates that, in CMS' view, calibrating the model using MA encounter data would trigger this provision, removing the requirement.

  CMS might be able to continue to apply a coding pattern adjustment under its general authority at section 1853(a)(1)(C)(i) of the Act, but this would present some novel challenges. First, CMS would likely need to be prepared to present—and defend—statistical analyses justifying its choice of coding pattern adjustment. That is not currently necessary since CMS has simply applied the statutory minimum adjustment. Second, CMS might face litigation arguing that coding pattern adjustments exceed its authority under section 1853(a)(1)(C)(i) of the Act since section 1853(a)(1)(C)(ii) of the Act already offers explicit instructions on when and how CMS should apply such adjustments.

- **Calibrating the model using MA encounter data would not necessarily produce better measures of relative risk:** The Advance Notice also argues that calibrating the model using the MA encounter data would improve the risk adjustment system's performance because "MA encounter data is likely a better predictor of relative costs in MA than FFS claims data." For two reasons, this is less clear than it may seem, coding issues aside.

  First, risk scores would ideally reflect the relative cost that an *efficient* plan would incur to cover different types of enrollees,[5] and it is unclear whether TM or MA spending patterns come closer to this ideal. A key difference between MA and TM is that MA plans have strong incentives to cater to beneficiary demands and the flexibility to do so. This can promote greater efficiency (e.g., by encouraging plans to root out low-value care and use the savings to finance higher-value benefits) or inhibit it (e.g., by encouraging plans to stint on care that is not salient to beneficiaries at the time they select a plan). Plans also have

[4] Medicare Payment Advisory Commission (MedPAC), "Medicare Payment Policy."

[5] Savannah L. Bergquist et al., "Intervening on the Data to Improve the Performance of Health Plan Payment Methods," Working Paper (National Bureau of Economic Research, April 2018), https://doi.org/10.3386/w24491.

incentives to avoid enrollees whose costs are not adequately offset by the risk adjustment system, which could cause plans to stint on care for some types of enrollees.

Second, it is unclear whether a model calibrated using the MA encounter data would even do a better job matching current relative costs in MA. MA plans make extensive use of non-fee-for-service payments.[6] Many such payments are likely missing from the encounter data. In other cases, these payments may be set in ways that obscure how resource use varies across different types of enrollees (consider, e.g., a capitation arrangement where the capitation amount reflects the average cost of a provider's patients). While TM also uses alternative payment models, these problems are likely less severe in TM since TM's largest alternative payment models operate on a fee-for-service chassis, with only year-end reconciliation payments occurring outside the ordinary claims-payment process.[7]

To be clear, these points do not necessarily imply that CMS should avoid using the encounter data for calibration. However, they do imply that CMS should proceed cautiously, at least if its goals are to mitigate the payment consequences of plans' coding efforts or improve its measures of relative risk. Namely, CMS should carefully assess whether calibrating the model using encounter data would actually advance these goals, not simply assume it would do so. It should also develop a viable plan for applying a coding pattern adjustment in the changed environment.

Thank you for the opportunity to comment on the Advance Notice. I hope that this information is helpful to you. If I can provide any additional information, I would be happy to do so.

Sincerely,

Matthew Fiedler
Joseph A. Pechman Senior Fellow in Economic Studies
Center on Health Policy
Economic Studies Program
The Brookings Institution

[6] Health Care Payment Learning and Action Network, "APM Measurement: Progress of Alternative Payment Models," November 14, 2024, https://hcp-lan.org/wp-content/uploads/2024/11/2024-HCPLAN-Methodology-Report-11-13.pdf.

[7] CMS also holds comprehensive data on its payments under alternative payment models. In principle, CMS could likely reflect at least some of these payments in the enrollee-level data used for model calibration.

Appendix: Calibration Data Source and the Appropriate Coding Pattern Adjustment

This appendix presents a model of how changing the data used to calibrate the risk score model would change the adjustment needed to account for coding differences between MA and TM.

Model setup

Consider a population of Medicare beneficiaries indexed by i who can choose between two coverage types $j \in \{\mathrm{TM}, \mathrm{MA}\}$. To focus on coding intensity issues, I assume that there is no selection into MA, so TM and MA beneficiaries are each representative of beneficiaries overall.

Suppose, for simplicity, that risk scores reflect the presence or absence of a single diagnosis. That is, risk scores take the form $R(D; \{\tau_d\}) = \tau_0(1-D) + \tau_1 D$, where $D \in \{0,1\}$ is the diagnosis indicator and τ_d are model coefficients. Let $D_i^j \in \{0,1\}$ capture whether that diagnosis is reported when beneficiary i selects coverage type j. I assume that coding intensity is higher in MA than in TM; formally, I assume that $D_i^{\mathrm{MA}} \geq D_i^{\mathrm{TM}}$, with $D_i^{\mathrm{MA}} > D_i^{\mathrm{TM}}$ for a non-zero share of beneficiaries.

Let Y_i denote the health care spending of beneficiary i; this amount is not indexed by j, reflecting my simplifying assumption that each beneficiary spends the same amount whether enrolled in TM or MA. Under these assumptions, the coefficients of a risk score model calibrated on data from coverage type $c \in \{\mathrm{TM,MA}\}$ are given by $\tau_d^c = \mathbb{E}[Y_i | D_i^c = d]$. I assume that having a diagnosis is always associated with higher expected spending; that is, $\tau_1^c > \tau_0^c$ for each data source c.

Effect of coding pattern differences on MA plan payments

I consider a simplified version of the MA payment system where the goal is to pay MA plans the expected cost of their enrollees; given the assumed absence of selection into MA, this is simply $\mathbb{E}[Y_i]$. In practice, however, payments are risk-adjusted, and payments prior to any coding pattern adjustment are given by $\mathbb{E}[Y_i] + \mathbb{E}\left[R\left(D_i^{\mathrm{MA}}; \{\tau_d^c\}\right)\right] - \mathbb{E}\left[R\left(D_i^{\mathrm{TM}}; \{\tau_d^c\}\right)\right]$.

The payment error when the risk score model is calibrated using data from coverage type c is

$$\mathbb{E}\left[R\left(D_i^{\mathrm{MA}}; \{\tau_d^c\}\right)\right] - \mathbb{E}\left[R\left(D_i^{\mathrm{TM}}; \{\tau_d^c\}\right)\right] = (\tau_1^c - \tau_0^c)\left(\mathbb{E}\left[D_i^{\mathrm{MA}}\right] - \mathbb{E}\left[D_i^{\mathrm{TM}}\right]\right).$$

That is, the payment error equals the incremental predicted cost of the diagnosis, $\tau_1^c - \tau_0^c$, multiplied by the difference in its reported prevalence between MA and TM, $\mathbb{E}\left[D_i^{\mathrm{MA}}\right] - \mathbb{E}\left[D_i^{\mathrm{TM}}\right]$.

Implications for the effects of changing the calibration data source

One immediate implication of this equation is that, regardless of the calibration data source, higher coding intensity in MA leads MA plans to be overpaid since, by assumption, $\mathbb{E}\left[D_i^{\mathrm{MA}}\right] > \mathbb{E}\left[D_i^{\mathrm{TM}}\right]$ and $\tau_1^c > \tau_0^c$. Thus, a coding pattern adjustment will always be needed to achieve the target payment level regardless of whether MA or TM data are used for calibration.

A second implication is that whether the overpayment rises or falls when calibration data source changes hinges on whether the incremental predicted cost of the diagnosis, $\tau_1^c - \tau_0^c$, grows or shrinks. This depends, in turn, on how the spending of the beneficiaries who get "upcoded" when they enroll in MA instead of TM compares to the spending of other Medicare beneficiaries.

It is instructive to consider two polar assumptions about these beneficiaries:

1. The first is that these beneficiaries have the same average spending as beneficiaries with $D_i^{\text{TM}} = 0$. This would be the case, for example, if MA plans randomly assign spurious diagnoses to people without a diagnosis. In this case, $\tau_0^{\text{MA}} = \tau_0^{\text{TM}}$ and $\tau_1^{\text{MA}} < \tau_1^{\text{TM}}$. Thus, the incremental predicted cost of the diagnosis shrinks, and the overpayment shrinks as well.

2. The second is that these beneficiaries have the same average spending as the beneficiaries with $D_i^{\text{TM}} = 1$. This would be the case if MA plans are identifying additional people who actually have the relevant condition and, moreover, who are identical to those already reported to have the condition in TM. In this case, $\tau_1^{\text{MA}} = \tau_1^{\text{TM}}$ and $\tau_0^{\text{MA}} < \tau_0^{\text{TM}}$. Thus, the incremental predicted cost of the diagnosis grows, and the overpayment grows as well.

As discussed in the main text, reality likely lies in between these extremes.