Center on Health Policy at BROOKINGS

REPORT | JANUARY 2025

How precise are the estimates of Traditional Medicare costs used to set Medicare Advantage benchmarks?

SAMANTHA CROW, MATTHEW FIEDLER, AND JULIA PARIS

The authors thank Richard Frank for helpful comments, Chloe Zilkha for research assistance, Rasa Siniakovas for editorial assistance, and Chris Miller for assistance with web posting. All errors are our own. This work was supported by a grant from Arnold Ventures.

The Brookings Institution is financed through the support of a diverse array of foundations, corporations, governments, individuals, as well as an endowment. A list of donors can be found in our annual reports, published online. The findings, interpretations, and conclusions in this report are solely those of its author(s) and are not influenced by any donation.

Executive Summary

The benchmarks used to determine payments to Medicare Advantage (MA) plans are based on the expected cost of covering a Traditional Medicare (TM) enrollee in a plan's local area. TM now enrolls less than half of all eligible Medicare beneficiaries, following the large and ongoing shift of beneficiaries into MA in recent years. This shift has spurred questions about whether TM enrollment is or soon will be too small to allow the Centers for Medicare and Medicaid Services (CMS) to construct reliable estimates of local TM costs. If so, this would undermine the logic of basing MA payments on estimates of local TM costs since these estimates would no longer be a reliable guide to the costs that MA enrollees would incur if they instead enrolled in TM.

In this paper, we estimate the amount of error in CMS' estimates of county-level TM costs attributable to the fact that CMS must rely on data for a finite population of TM enrollees. In addition to estimating sampling error, we estimate the non-sampling error introduced by the "credibility adjustments" that CMS applies to mitigate sampling error; these adjustments blend estimates for counties with fewer than 1,000 TM enrollees with estimates for the counties' regions and introduce error if the affected counties differ from their regions. We do not consider other reasons that observed TM spending may be a poor guide to what MA enrollees would cost if enrolled in TM, notably the possibility that TM and MA enrollees systematically differ from each other; these related, but distinct, issues have been studied in prior research.

We have four main findings:

- **The absolute number of TM enrollees, the main determinant of the amount of sampling error in CMS' estimates of county TM costs, is projected to be roughly stable over the coming decade.** Both the Medicare Trustees and the Congressional Budget Office expect declines in TM's market share to be offset by increases in the total number of Medicare beneficiaries over the next ten years. These projections do assume that TM's market share will decline more slowly over the next ten years than it did over the past decade. Nevertheless, they suggest that statistical challenges posed by low TM enrollment may not meaningfully worsen in the near-term.

- **Sampling error in CMS' estimates of TM costs is currently modest and will likely remain so even if TM's market share falls faster than expected.** We estimate that, in 2022, the median Medicare beneficiary lived in a county where the standard error of CMS' estimate of county TM costs was 0.6% of the underlying estimate. The 95th percentile beneficiary lived in a county where this value was 1.9%. Even in a scenario where TM's market share continues to fall rapidly, the median standard error would be 0.7% of the underlying estimate and the 95th percentile value would be 2.4% in 2033.

- **The credibility adjustments applied in counties with low TM enrollment introduce considerable non-sampling error, but the affected counties contain a small fraction of beneficiaries.** For counties subject to a credibility adjustment, we estimate that the total root mean squared error of CMS' estimates of county TM costs—accounting for both sampling error and the non-sampling error introduced by the credibility adjustment process—averaged 6.5% in 2022. By contrast, this estimate was just 3.5% when accounting for sampling error alone. However, only 1.2% of beneficiaries lived in counties subject to a credibility adjustment in 2022, and only 2.2% of beneficiaries will live in such counties by 2033 even in the lowest-enrollment scenario we consider.

- **Changes to the credibility adjustment process could increase the accuracy of CMS' estimates for counties with very low TM enrollment.** We estimate that the current credibility adjustment process *increases* the root mean squared error of CMS' estimates for many affected counties relative to applying no adjustment. CMS could improve the accuracy of its estimates for these counties by changing how it ensures that the credibility adjustment is budget neutral and how it determines the weight placed on the regional estimate versus the county's own estimate.

We conclude that CMS can construct fairly precise estimates of local TM costs at both current TM enrollment levels and those likely to prevail over the next decade. Thus, concerns about the precision of CMS' estimates offer, at best, a weak basis for proposals to "decouple" MA payment from local TM costs, although there may be other rationales for such changes (e.g., concerns about the performance of the MA risk adjustment system). Our results do, however, suggest that CMS should refine how it estimates TM costs in counties with very low TM enrollment.

Introduction

By statute, the benchmarks used to determine payments to Medicare Advantage (MA) plans are based on the expected cost of covering an enrollee under Traditional Medicare (TM) in a plan's local area. TM now enrolls slightly less than half of all eligible Medicare beneficiaries, following a large shift of beneficiaries into MA in recent years (Freed et al. 2024). With TM's market share projected to continue declining (Trustees 2024; CBO 2024), some observers have questioned whether it will remain advisable to base MA benchmarks on local TM costs (see, e.g., McWilliams 2022; Jacobson and Blumenthal 2022; MedPAC 2023a; Lieberman, Ginsburg, and Valdez 2023).

There are two concerns in this vein. The first is that the absolute number of TM enrollees in many counties, the geographic unit for which CMS estimates local TM costs, may become very small. This could inject large sampling errors into CMS' estimates. It could also force CMS to rely heavily on the "credibility adjustments" it uses to mitigate sampling error; these adjustments blend estimates for counties with fewer than 1,000 enrollees with estimates for the counties' regions and introduce non-sampling error if the affected counties differ from their regions. The second—and distinct—concern is that, as TM shrinks, enrollees who remain in TM may increasingly differ from MA enrollees (in ways that the MA risk adjustment system does not capture). Either concern could make observed TM spending a poor guide to what MA enrollees would cost if they were instead enrolled in TM and, thus, undermine the logic of basing MA payments on TM costs.

The goal of this paper is to assess the first concern. To that end, we estimate the amount of error in CMS' estimates of local TM costs due to the fact that CMS must rely on data for a finite population of TM enrollees. We examine both sampling error and the non-sampling error introduced by credibility adjustments. To shed light on how these sources of error may evolve over time, we construct estimates for the actual enrollment levels reflected in the CMS estimates used to determine MA payments for the 2022 plan year and two future enrollment scenarios.

The paper proceeds as follows. We first describe the MA payment system and how CMS estimates county TM costs. We then examine how TM enrollment is likely to evolve in the coming years. Next, we estimate the sampling error in CMS' estimates in our three enrollment scenarios, as well as the non-sampling error introduced by the credibility adjustment process. We close by considering the implications of our results for proposals to "decouple" MA benchmarks from local TM costs, as well as ways CMS could improve the accuracy of its estimates of local TM costs.

Background on MA Payment and CMS' Estimates of County TM Costs

MA plans receive monthly risk-adjusted capitation payments for each person they enroll. Capitation payments to MA plans generally equal the bid submitted by the plan plus a "rebate" equal to a percentage of the difference between the bid and a benchmark established by CMS.[1]

Benchmarks are generally established at the county level as a percentage of the cost CMS expects to incur to cover an enrollee of average risk under TM in that county and plan year.[2] That percentage varies based on a plan's location and quality performance. In 2024, benchmarks

---

[1] In the rare cases where a plan's bid exceeds the benchmark, the federal government pays the benchmark amount, and the beneficiary is required to cover the remainder of the bid through a supplemental premium. For a fuller overview of the MA payment system, see MedPAC (2023b).

[2] There are exceptions to this general rule: benchmarks for enrollees with end-stage renal disease are established at the state level; in some counties, benchmarks are subject to a statutory upper limit; and regional preferred provider organization plans have different rules. For details, see MedPAC (2023b).

nominally average 108% of TM costs, although benchmarks are functionally much higher due to shortcomings of the MA risk adjustment system, including higher diagnosis coding intensity in MA and favorable selection into MA (MedPAC 2024).

The focus of this paper is the reliability of the underlying estimates of county TM costs. CMS publishes these estimates annually as part of the MA rate setting process. Here and throughout, we focus on estimates CMS published for the 2022 plan year and the methods underlying those estimates. This was the last year that did not incorporate historical data influenced by the COVID-19 pandemic, which allows us to avoid any pandemic-related complications. However, the process used to compute these estimates has changed little since 2022, so our main findings should be applicable to later years.

CMS' methodology for the 2022 plan year consisted of the following four main steps:[3]

1. *Tabulate historical TM spending, enrollment, and risk scores by county:* CMS tabulated Part A and Part B (non-hospice) fee-for-service claims payments, Part A and Part B enrollment, and average CMS-HCC risk scores for TM enrollees without end-stage renal disease for the most recent five years with available data, in this case, 2015 through 2019.[4]
2. *Compute county risk-adjusted TM costs relative to the national average:* CMS next computed per member per month Part A and Part B spending for each county and historical year, summed them to obtain a combined Part A and B cost, normalized the results by dividing by their national enrollment-weighted average, and averaged the normalized amounts across the five historical years. To account for difference in enrollee risk across counties, CMS then divided these amounts by the simple average of the average risk scores for the five historical years before normalizing the resulting amounts a second time by dividing by their national enrollment-weighted average.
3. *Multiply by a projection of national average per enrollee TM costs in 2022:* CMS multiplied the relative cost calculated for each county in step 2 by CMS' projection of national average per enrollee Part A and B spending for TM enrollees without end-stage renal disease for 2022. This calculation yielded a provisional estimate of the expected TM cost in the county for 2022.
4. *Apply a "credibility adjustment" for counties with very low TM enrollment:* After calculating these provisional estimates, CMS applied a credibility adjustment to counties with fewer than 1,000 TM Part B enrollees on average across the five historical years, an enrollment measure we refer to as the county's "credibility enrollment" in what follows. For each such county, CMS calculated a blended estimate that was a weighted average of the county's own provisional estimate and the provisional estimate for the county's region, defined as the county's core-based statistical area (CBSA) or, for non-CBSA counties, all non-CBSA

[3] This description omits several minor steps in CMS' calculations. Notably, CMS: (1) adjusts TM claims experience for price changes and to incorporate certain non-fee-for-service payments; (2) adjusts its estimates for Puerto Rico to reflect certain unique circumstances; (3) adjusts its estimates to remove direct graduate medical education costs, indirect medical education costs, and kidney acquisition costs; and (4) makes an adjustment intended to remove the effect of the presence of enrollees who are eligible for coverage via the Department of Veterans Affairs or the Department of Defense. For full details, see CMS (2020; 2021a; 2021b).

[4] The average risk scores tabulated by CMS appear to exclude months where an enrollee lacked Part A or Part B coverage, and CMS' documentation indicates that they exclude hospice enrollment months.

counties in the state. In this blend, the county's own estimate received a weight $\sqrt{M_c / 1000}$, where M_c is the county's credibility enrollment, with the remaining weight assigned to the regional estimate. CMS then applied a "budget neutrality" adjustment that increased or decreased the blended estimates by the state-specific percentage required to ensure that the credibility adjustment process neither increased nor decreased CMS' estimates of TM costs, on average, across the affected counties in a state.

Our goal in this paper is to estimate the amount of error in CMS' estimates due to the fact that it must rely on data for a finite population of TM enrollees. To be precise, we are interested in error relative to the estimates that CMS would hypothetically obtain if it held data for a arbitrarily large population of TM enrollees drawn from the same distribution as actual TM enrollees.

In practice, CMS' estimates will deviate from this target for two reasons. First, due to random chance, the actual population of TM enrollees will not be perfectly representative of this hypothetical larger population, generating sampling error in the tabulations described in step 1 and, in turn, in CMS' estimates of county TM costs. Second, with a finite TM enrollee population, CMS will apply credibility adjustments in some counties, as described in step 4. These adjustments will introduce non-sampling error to the extent that underlying TM costs differ between the affected counties and their regions.

This paper focuses on the error in CMS' estimates due to population size constraints because it has not, to our knowledge, been studied elsewhere. But this is not the only potential source of error in MA payment. Indeed, as noted above, there is substantial evidence that the MA risk adjustment system is failing to fully adjust for differences in claims risk between the average TM enrollee (which is what the CMS estimates of county TM cost are supposed to reflect) and MA enrollees; this can also undermine the underlying statutory goal of basing payments to MA plans on what it would cost to cover MA enrollees through TM.

Medium-Run Projections of TM Enrollment

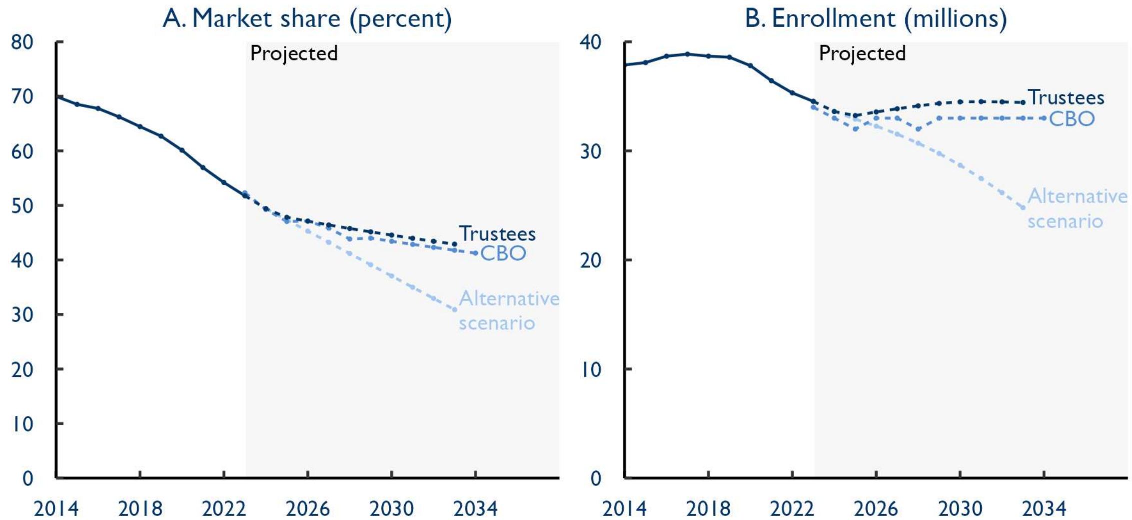
Both the amount of sampling error in CMS' estimates and the scope for credibility adjustments to introduce non-sampling error depend on the number of people enrolled in TM. Thus, we begin our analysis by examining national projections of TM enrollment for the coming years.

Figure 1 shows recent TM enrollment projections by the Medicare Trustees and the Congressional Budget Office (CBO). Each projects that TM's *share* of total Medicare enrollment will continue to decline over the coming decade. The Medicare Trustees (2024) project that TM's market share will decline from 49% in 2024 to 43% in 2033, while CBO (2024) shows a slightly larger decline.

However, both forecasts show little change in the *absolute number* of TM enrollees over this period because falling TM market share is offset by growth in the total number of Medicare beneficiaries driven by population aging. Thus, if these forecasts are borne out, the size of the sampling and non-sampling errors of interest here might not change much over the next decade.

It is notable, however, that both the Trustees and CBO are predicting that TM's market share will decline more slowly over the coming decade than in the recent past. Thus, Figure 1 also presents an alternative scenario in which TM's market share declines at the same rate (measured in percentage points per year) from 2024 onward as it did from 2014 through 2024. In this scenario, the absolute number of TM enrollees does decline markedly over the decade, from 34 million in

Figure 1. Traditional Medicare Enrollment, Historical and Projected



Notes: Historical data are from the Medicare Trustees (2024). The Trustees and CBO scenarios reflect projections from the Medicare Trustees (2024) and CBO (2024), respectively. The alternative scenario assumes that TM's market share matches the Medicare Trustees in 2024, then declines by the same number of percentage points per year as occured from 2014 through 2024, as estimated using the Trustees' historical estimates and their projections for 2024.

2024 to 25 million in 2033, a decline of 26%. In this alternative scenario, the size of the relevant errors in CMS' estimates likely would grow over time.

Estimating Sampling Error in CMS' TM Cost Estimates

We now estimate the sampling error in CMS' estimates of expected TM costs under various TM enrollment scenarios. It is difficult to derive an analytic formula for this sampling error because CMS' estimates are a complicated non-linear function of the input data. Thus, we use a bootstrap approach. This section describes our bootstrap method and then presents our results.

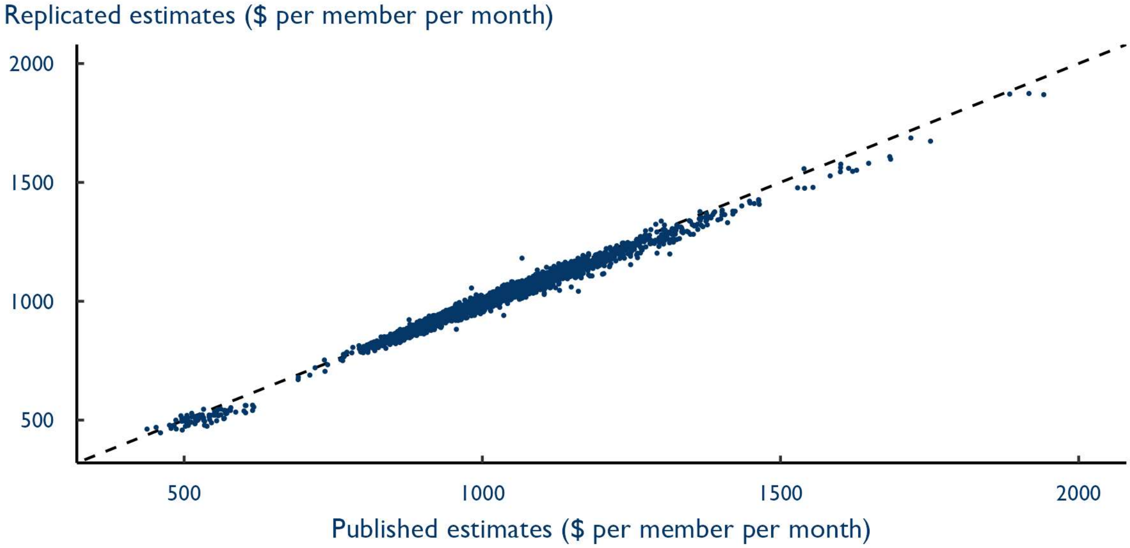
Bootstrap method

We rely on Medicare claims and enrollment data accessed via the Virtual Research Data Center (VRDC). As a first step, we use the VRDC data to replicate the dataset that CMS used to create its tabulations of county-level TM spending, enrollment, and risk scores. We then reimplement the calculations that CMS uses to transform those tabulations into estimates of county TM costs.[5]

To assess the accuracy of our replication, we compare our estimates of county TM costs to CMS' published estimates. They align fairly closely, albeit not perfectly, as shown in Figure 2. The difference between our estimates and CMS' published estimates (expressed as a percentage of the CMS estimate) has a mean of -1.0% and a standard deviation of 1.8%, and the correlation between the two sets of estimates is 0.99. The discrepancies that do exist reflect differences between our tabulations of TM claims data and those created by CMS, not differences in how we

[5] We do not fully reimplement the minor steps in CMS' calculations that were described in footnote 3, some of which rely on input data we lack access to. Rather, we take advantage of the fact that CMS publishes a spreadsheet that reports the intermediate results of every step in the calculations it uses to transform its tabulations of TM claims data into its final estimates. We calculate the county-specific proportional change that occurs in each relevant step of CMS' calculations and then apply the same proportional change in our calculations.

Figure 2. Replicated versus Published Estimates of County TM Costs



Note: Published estimates of county TM costs are those produced by CMS for the 2022 plan year. Replicated estimates are calculated as described in the text. The dashed line depicts where the replicated and published estimates are equal.

transform those tabulations into estimates of county TM costs.[6] Those discrepancies may, in turn, reflect differences between our interpretation of CMS' methodology for tabulating TM claims and CMS' actual methodology, as well as differences between the claims data available on the VRDC and the internal databases used by CMS. (To ensure internal consistency, all subsequent analyses rely solely on our replicated estimates, not CMS' published estimates.)

To estimate the sampling distribution of CMS' estimates of county TM costs, we take advantage of the fact that they can be written as $g(\{\bar{X}_c\},\{N_c\})$, where each $\bar{X}_c = [1/N_c]\sum_{i:C_i=c} X_i$ is the sample mean of a vector of beneficiary-level variables X_i across the set of beneficiaries i with county of residence $C_i = c$, and N_c is the number of such beneficiaries.[7,8] Each component of X_i captures a single aspect of the beneficiary's TM experience (e.g., spending or enrollment months) for one year of the 2015-2019 period.[9] In terms of the description of CMS' methods in the last section, the amounts $\{\bar{X}_c\}$ and $\{N_c\}$ correspond to the tabulations under step 1, while the function g corresponds to the calculations described in steps 2-4.

---

[6] In particular, when we input CMS' TM claims tabulations into our code, we obtain final estimates of county TM costs that are essentially identical to CMS', with minor exceptions that appear to reflect CMS' suppression of data for several small counties and an apparent error in CMS' calculations for a single county.

[7] The beneficiary counts N_c enter this function because some steps in CMS calculations involve calculating nationwide enrollment-weighted averages and because of the credibility adjustment.

[8] If a beneficiary appears in multiple counties during the 2015-2019 period, we treat each beneficiary-county pair as a distinct beneficiary.

[9] For most of the tabulations of TM experience that enter CMS' calculations, it is obvious how to define the corresponding components of X_i. One exception is the average risk score among TM enrollees in each county-year cell. This average cannot be expressed as a simple mean of beneficiary-level values because it is weighted by the number of months the beneficiary was in the risk score universe during the year. However, it can be expressed as the *ratio* of the means of two beneficiary-level characteristics: (1) the number of "risk score months," meaning the sum of the risk scores that applied to the beneficiary in each month the beneficiary was in the universe for which CMS tabulates risk scores; and (2) the number of months the beneficiary was in that universe.

We assume that each $\bar{X}_c$ is normally distributed with mean $\mu_c = \mathbb{E}[X_i | C_i = c]$ and covariance matrix Σ_c / N_c, with $\Sigma_c = \text{Var}(X_i | C_i = c)$, which will be approximately true when N_c is large.[10] Importantly, this structure allows for arbitrary within-beneficiary correlations across different aspects of beneficiary experience (e.g., spending and enrollment duration) and across years. Letting $\hat{\mu}_c$ and $\hat{\Sigma}_c$ denote the sample counterparts of μ_c and Σ_c estimated in replicated sample, we draw 1,000 replicates of each vector $\bar{X}_c$ from the distribution $N(\hat{\mu}_c, \hat{\Sigma}_c / N_c)$, denoting replicate r by $\bar{X}_c^r$. We use the empirical distribution of $g(\{\bar{X}_c^r\}, \{N_c\})$ to approximate the sampling distribution of the vector of CMS' TM cost estimates.

This approach generates an estimate of the sampling distribution of CMS' estimates of TM costs at the *actual* TM enrollment levels reflected in the 2015-2019 data used in CMS' calculations for the 2022 plan year. However, we also consider two other TM enrollment scenarios derived from the enrollment projections in Figure 1. In the first, we assume that TM enrollment matches the Trustees' projection for 2033, which is 11% below the 2015-2019 average. In the second, we assume that TM enrollment matches the 2033 projection from the "alternative" scenario, which is 36% below the 2015-2019 average.

To estimate the sampling distribution in these scenarios, we assume that beneficiaries are still drawn from the same underlying distribution.[11] Based on this assumption, we draw two additional sets of replicate means $\bar{X}_c^r(\alpha)$ from the distribution $N(\hat{\mu}_c, \hat{\Sigma}_c / [\alpha N_c])$, where α is the ratio of TM enrollment in the scenario in question to 2015-2019 average enrollment.[12] We then use the empirical distribution of $g(\{\bar{X}_c^r(\alpha)\}, \{\alpha N_c\})$ to approximate the sampling distribution of interest.

Results

We now present standard errors for CMS' estimates of county TM costs, calculated as the sample standard deviation of the replicate estimates. To ease interpretation, we rescale each standard error by dividing by the relevant county TM cost estimate and present the result as a percentage.

Figure 3 depicts a histogram that summarizes the distribution of scaled standard errors, while Panel A of Table 1 reports summary statistics.[13] In each, we weight counties by total Medicare enrollment (specifically, the number of beneficiary-years of simultaneous Part A and B enrollment in either TM or MA by beneficiaries without end-stage renal disease in the county during 2022).

---

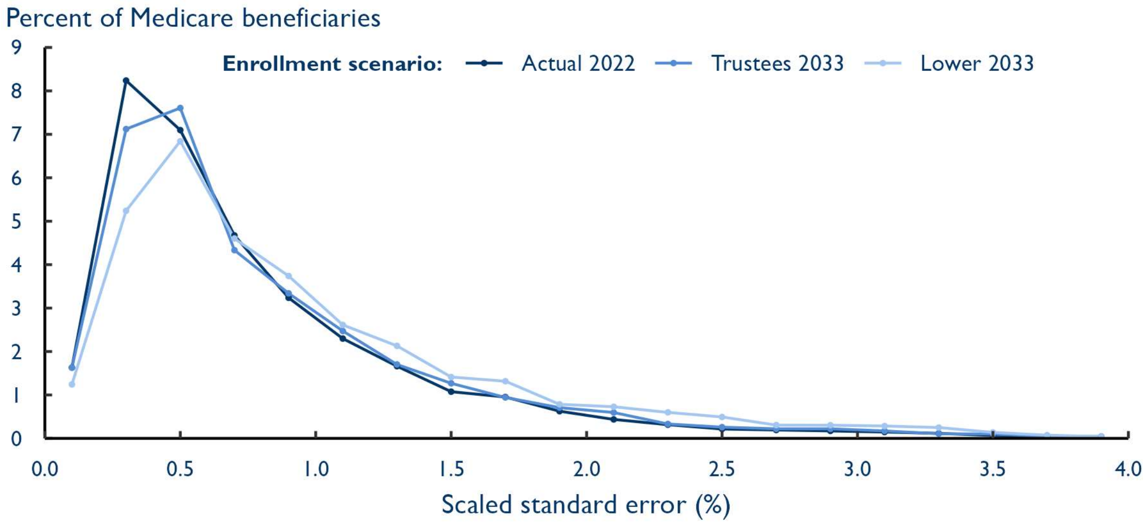
[10] We experimented with a non-parametric bootstrap, which may be more robust to violations of this normality assumption in smaller counties. In practice, the results were very similar.

[11] This assumption is unlikely to hold exactly in a world with lower TM enrollment; in particular, it implies that reduced TM enrollment occurs entirely through a reduction in how many people *ever* enroll in TM, with no change in the duration or time pattern of TM enrollment among those with some enrollment. But there is little reason to expect this to substantially bias our estimates of the measures of sampling error of primary interest in this paper, especially since we scale our estimates of sampling variation by the means of the relevant distributions.

[12] In practice, we simply rescale the initial set of replicates using $\bar{X}_c^r(\alpha) = \hat{\mu}_c + \alpha^{-1/2}[\bar{X}_c^r - \hat{\mu}_c]$. This generates replicates with the desired distribution but allows for less noisy comparisons across enrollment scenarios because all three sets of replicates reflect the same underlying random draws.

[13] In some instances, CMS pools data for a group of counties and reports a single estimate for those counties. We follow CMS in this regard but present results for only one county in each group.

Figure 3. Distribution of Scaled Standard Errors of TM Cost Estimates



Note: For each enrollment scenario, the figure plots a histogram of the distribution of bootstrapped standard errors for CMS' estimates of county TM costs, divided by the underlying TM cost estimates. The histogram bin width is 0.2%, and each point is plotted at the bin midpoint. Counties are weighted by Medicare enrollment. See text for details.

Table 1. Distribution of Scaled Standard Errors for TM Cost Estimates

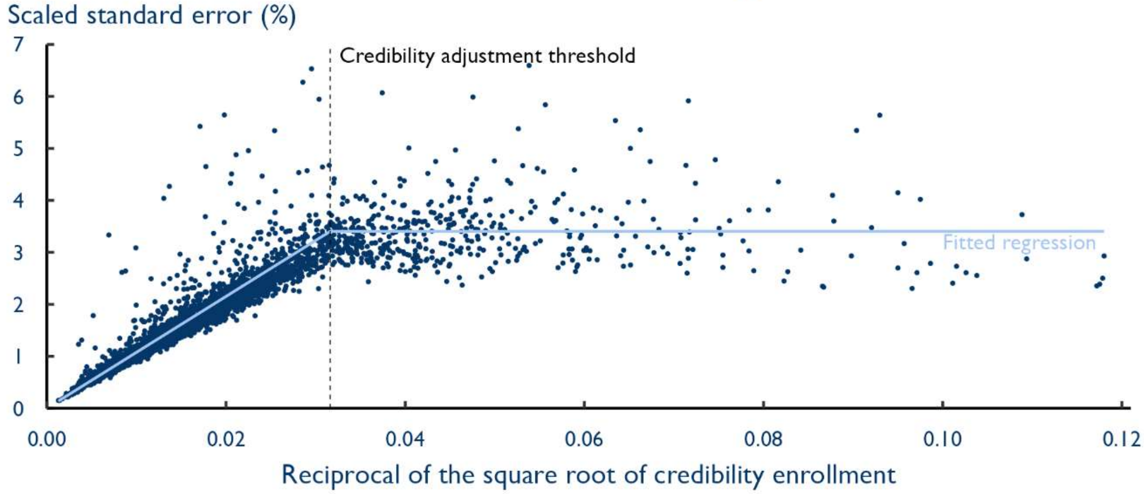
		Percentiles of Distribution (%)			
Scenario	**Mean (%)**	25th	50th	75th	95th
Panel A. Base estimates					
Actual 2022	0.8	0.4	0.6	1.0	1.9
Trustees 2033	0.8	0.4	0.6	1.0	2.0
Lower 2033	1.0	0.5	0.7	1.2	2.4
Panel B. Fitted values from regression of raw estimates on county enrollment					
Actual 2022	0.8	0.4	0.6	0.9	1.8
Trustees 2033	0.8	0.4	0.6	1.0	1.9
Lower 2033	0.9	0.4	0.6	1.1	2.1

Note: For each enrollment scenario, Panel A presents summary statistics for the distribution of bootstrapped standard errors for CMS' estimates of county TM costs, divided by the underlying TM cost estimates. Panel B presents the same summary statistics calculated using the fitted values from a regression of the scaled standard errors on county TM enrollment. Counties are weighted by Medicare enrollment. See text for details.

The median scaled standard error of actual estimates for the 2022 plan year is 0.6%, while the 95th percentile scaled standard error is 1.9%. For the two scenarios with lower enrollment, the estimates are, naturally, higher. For the lowest-enrollment scenario that we consider, the median scaled standard error is 0.7%, while the 95th percentile scaled standard error is 2.4%.

Figure 4 shows that cross-county variation in the scaled standard errors is mostly explained by differences in the level of TM enrollment (measured here by the enrollment measure used for credibility adjustments). The relationship between the scaled standard errors and TM enrollment takes roughly the form expected on theoretical grounds. At high enrollment levels, the scaled standard errors vary linearly with the reciprocal of the square root of enrollment, reflecting the usual relationship between sampling error and sample size. Then, once enrollment falls low enough that the county becomes subject to a credibility adjustment, the scaled standard errors cease to vary with enrollment; this is because the increase in sampling error due to reduced

Figure 4. Scaled Standard Errors versus Credibility Enrollment



Note: Each point is a county; the x-value reflects the reciprocal of the square root of credibility enrollment, and the y-value reflects the bootstrapped standard errors for CMS' estimates of county TM costs, divided by the underlying TM cost estimates. The line depicts a fitted regression that is linear in the reciprocal of the square root of credibility enrollment for enrollment above 1000 and constant below that. 16 county estimates fall outside the range of the plot.

sample size is precisely offset by a reduction in the weight that CMS places on the county's own experience relative to experience for the county's region.[14] A regression using this functional form, which is depicted by the light blue line in Figure 4, has an R^2 of 0.93.[15]

One implication of the tight relationship between the scaled standard errors and TM enrollment is that sampling error in the standard errors themselves has little effect on our overall findings. Indeed, Panel B of Table 1 replicates our main result, except with the raw scaled standard error for each county replaced by the predicted value from the regression line depicted in Figure 4. This approach can be interpreted as "turning off" variation in our standard errors due to *either* sampling variation or true variation in enrollee characteristics across counties. As such, it likely overstates the effect of removing sampling variation alone. Nevertheless, our results are little changed.

Estimating Non-Sampling Error Due to Credibility Adjustments

The preceding results show that the credibility adjustment process is effective at limiting sampling error in CMS' estimates for counties with very low TM enrollment. However, these adjustments could also introduce non-sampling error if expected TM costs differ between the counties subject to the adjustment and the regions used for blending. This section assesses how often credibility adjustments are applied, the overall accuracy of CMS' estimates for the affected counties taking account of both sampling error and this non-sampling error, and whether CMS' current methods optimally balance the two sources of error.

---

[14] To see this formally, consider a simplified version of CMS' methodology where the estimate of interest for each county c is the beneficiary-level mean $\bar{Y}_c$ when county enrollment $m_c \geq 1000$ and a blended estimate $w\bar{Y}_c + (1-w)\bar{Y}_c^R$ with $w = \sqrt{m_c \,/\, 1000}$ when $m_c < 1000$, where $\bar{Y}_c^R$ is the regional estimate used for blending. Assume that the regional estimate $\bar{Y}_c^R$ is independent of the county's own estimate $\bar{Y}_c$. For $m_c \geq 1000$, the variance takes the usual form $\sigma^2 \,/\, m_c$. For $m_c < 1000$, the variance is $\sigma^2/1000 + (1-w)^2\text{Var}(\bar{Y}_c^R)$. If $\text{Var}(\bar{Y}_c^R)$ is small relative to $\sigma^2 \,/\, 1000$, which is likely in practice since $\bar{Y}_c^R$ reflects average spending over a relatively broad region, then the variance of the blended estimate will be approximately constant for $m_c < 1000$.

[15] Formally, the figure depicts a regression of the form $S_c = \beta\max\{E_c, 1000\} + \epsilon_c$, where S_c is the scaled standard error estimate for county c, E_c is credibility enrollment for that county, and ϵ_c is the error term.

Table 2: Applicability of Credibility Adjustment by Enrollment Scenario

Enrollment scenario	Number of counties	Share of beneficiaries (%)	Mean weight assigned to regional estimate (%)
Actual 2022	486	1.2	27
Trustees 2033	551	1.4	26
Lower 2033	776	2.2	27

Note: The final column reflects only counties subject to a credibility adjustment and is weighted by the number of Medicare beneficiaries in the county. The number of beneficiaries in a county reflects the number of person-years of Medicare enrollment by people without end-stage renal disease and with both Part A and Part B coverage in 2022.

Prevalence of credibility adjustments

A first question is how often CMS would apply credibility adjustments in different enrollment scenarios. Table 2 reports how many counties were subject to an adjustment for the 2022 plan year, as well as how many counties would have been subject to an adjustment if enrollment were proportionally lower in accordance with the two 2033 enrollment scenarios considered previously.

Across all three scenarios, only a small fraction of Medicare beneficiaries is in counties that are subject to a credibility adjustment: 1.2% of beneficiaries for the 2022 plan year and 2.2% in the lowest enrollment scenario we consider. For counties where a credibility adjustment is applied, around one-quarter of the weight is assigned to the regional estimate rather than the county's own estimate, on average.

Estimating error inclusive of non-sampling error

A second question is how much non-sampling error credibility adjustments introduce when applied. To answer this question, we lay out a model of how CMS blends county and regional estimates (leaving aside, for now, the budget neutrality step in the credibility adjustment process).

We let Y_c denote the estimate of expected TM cost for county c before any credibility adjustment and let Y_c^R denote the corresponding estimate for the county's region. The county's own estimate is assigned a weight $w(M_c)$, where M_c is credibility enrollment in county c, resulting in a blended estimate $\tilde{Y}_c = w(M_c)Y_c + [1 - w(M_c)]Y_c^R$.[16] We assume that Y_c and Y_c^R are independent conditional on a vector of distributional parameters θ_c that includes the estimates' means μ_c and μ_c^R as well as M_c. This approach assumes that the sampling errors in the two estimates are independent. We then view each parameter vector θ_c as having been drawn from some higher-level distribution.

CMS' goal is to estimate the true mean μ_c for each county c. The mean squared error of the blended estimate with respect to this parameter (at an enrollment level M_c) can be written as

$$\mathbb{E}\left[\left(\tilde{Y}_c - \mu_c\right)^2 \mid M_c\right] = \underbrace{w(M_c)^2\mathbb{E}[\text{Var}(Y_c|\theta_c) \mid M_c] + \left(1 - w(M_c)\right)^2\mathbb{E}[\text{Var}(Y_c^R|\theta_c) | M_C]}_{\text{sampling error terms}}$$
$$+ \underbrace{\left(1 - w(M_c)\right)^2\mathbb{E}[(\mu_c^R - \mu_c)^2 | M_c]}_{\text{bias term}}. \quad (1)$$

The first two terms reflect the sampling error in Y_c and Y_c^R, respectively. The third term reflects the fact that the blended estimate is generally a biased estimator of the true county mean μ_c since,

[16] CMS implements the credibility adjustment using an estimate of Y_c^R that encompasses data from county c itself in addition to the other counties in its region, whereas our definition excludes the county's own data from Y_c^R. CMS' procedure can be viewed as reflecting a slightly different choice of weights.

in general $\mu_c^R \neq \mu_c$; note that the bias in the blended estimate is $\mathbb{E}[\tilde{Y}_c|\theta_c] - \mu_c = [1 - w(M_c)](\mu_c^R - \mu_c)$. This term shows why the credibility adjustment process can introduce non-sampling error.

We estimate the overall mean squared error of the blended estimate by estimating each term on the right-hand side of equation (1). In producing these estimates, we focus on the error in the estimates CMS produced for the 2022 plan year and, correspondingly, on the set of counties subject to a credibility adjustment for that year (that is, the counties in the first row of Table 2).[17]

To estimate the two sampling error terms, we rely on sampling error estimates obtained using the same bootstrap procedure as above (but applied to CMS' pre-credibility-adjustment estimates, not its final estimates). For $\mathbb{E}[\text{Var}(Y_c|\theta_c) \mid M_c]$ term, which captures sampling error in the county's own estimate, we assume a functional form $\mathbb{E}[\text{Var}(Y_c|\theta_c) \mid M_c] = \alpha / M_c$ and estimate the parameter α by an ordinary least squares regression of $\text{Var}(Y_c|\theta_c)$ on $1 / M_c$. For the $\mathbb{E}[\text{Var}(Y_c^R|\theta_c) \mid M_C]$ term, which captures sampling error in the regional estimate, we assume that the variance of the regional estimate $\text{Var}(Y_c^R|\theta_c)$ is independent of the enrollment level M_c so that $\mathbb{E}[\text{Var}(Y_c^R|\theta_c) \mid M_C] = \mathbb{E}[\text{Var}(Y_c^R|\theta_c)]$.[18] We then estimate $\mathbb{E}[\text{Var}(Y_c^R|\theta_c)]$ by taking the simple average of the bootstrapped estimates of $\text{Var}(Y_c^R|\theta_c)$.

Estimating the bias term is more involved. We assume that the county-region difference $\mu_c^R - \mu_c$ is independent of the enrollment level M_c so that $\mathbb{E}[(\mu_c^R - \mu_c)^2|M_c] = \mathbb{E}[(\mu_c^R - \mu_c)^2]$.[19] To estimate the right-hand side of this equation, we define $\Delta = \mathbb{E}[\mu_c^R - \mu_c]$ and $\tau^2 = \text{Var}[\mu_c^R - \mu_c]$ and note that $\mathbb{E}[(\mu_c^R - \mu_c)^2] = \Delta^2 + \tau^2$. The parameter Δ is the average difference between a county and its region, while the parameter τ^2 captures how much that difference varies from county to county.

To estimate Δ, we use the natural estimator:

$$\hat{\Delta} = [1/|\mathcal{C}|] \sum_{c \in \mathcal{C}} Y_c^R - Y_c\,, \tag{2}$$

where $\mathcal{C}$ is the set of counties subject to a credibility adjustment. To estimate τ^2, we observe that

$$\tau^2 = \text{Var}(Y_c^R - Y_c) - \mathbb{E}[\text{Var}(Y_c|\theta_c)] - \mathbb{E}[\text{Var}(Y_c^R|\theta_c)] \tag{3}$$

and obtain an estimator $\hat{\tau}^2$ by replacing each term on the right-hand side by an estimated counterpart. For the first term, we use the usual estimator of the sample variance, calculated in the set of counties subject to a credibility adjustment. For the second two terms, we take the simple average of our bootstrapped variance estimates for that set of counties.

The first row of Table 3 reports the resulting estimates of Δ and τ. The true underlying TM cost is \$59 per enrollee per month lower in the affected counties than in their regions, 5.7% of the mean

---

[17] We also produced estimates for the broader set of counties that may become subject to credibility adjustments in the future if TM enrollment continues to fall. Results were generally quite similar.

[18] We examined this assumption empirically by regressing our bootstrapped estimates of $\text{Var}(Y_c^R|\theta_c)$ on credibility enrollment M_c. These results suggested that the variance of the regional estimate may be modestly larger in low-enrollment counties. However, the variance of the regional estimate is so small in magnitude relative to the other terms in equation (1) that accounting for this relationship would have little effect on our results.

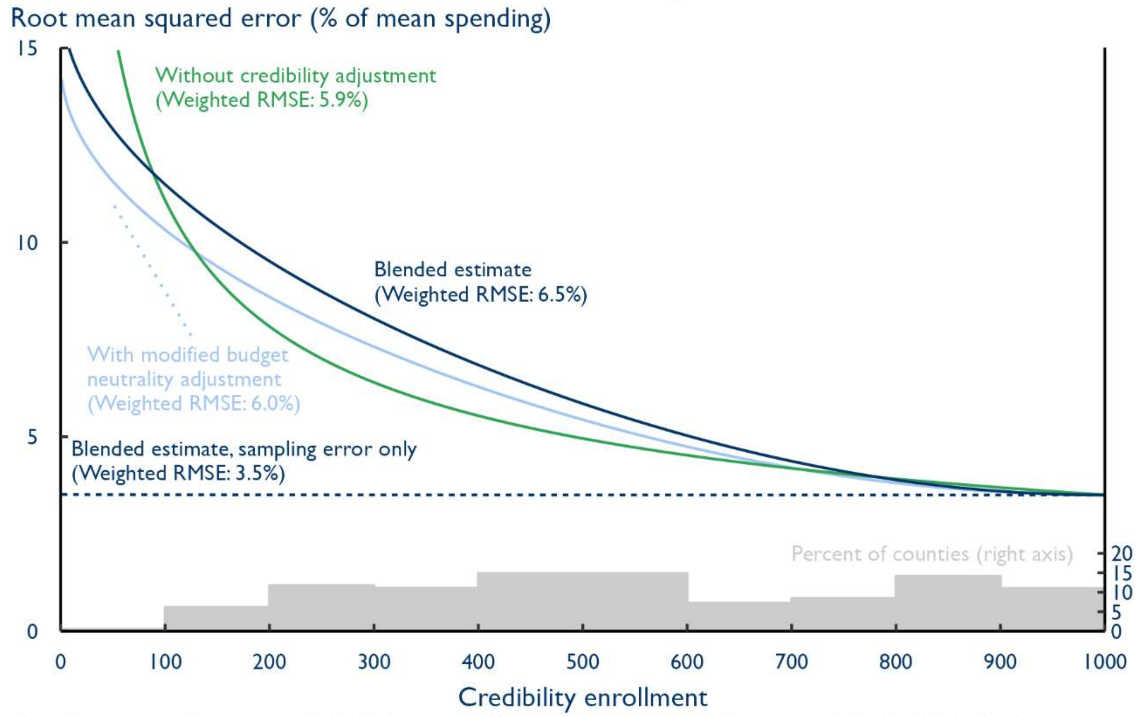
[19] We examined this assumption empirically by regressing $(Y_c^R - Y_c)^2$, less our estimates of the sampling variance of the county and regional estimates, on credibility enrollment M_c. The results suggest that there is little relationship between the squared difference $(\mu_c^R - \mu_c)^2$ and enrollment.

Table 3. Estimates of Underlying County-Region TM Cost Differences

Version of county-region differences	Dollars per enrollee per month		Percent of mean pre-credibility-adjustment estimate	
	Mean (Δ)	SD (τ)	Mean (Δ)	SD (τ)
Unadjusted	-59	156	-5.7	15.0
Adjusted to remove state average differences	0	148	0.0	14.2

Note: The table presents estimates of the mean underlying county-region cost difference (Δ) and the standard deviation of those differences (τ) estimated using the procedure outlined in the text. The "unadjusted" row presents parameters for the raw county-region differences, while the "adjusted to remove state average differences" row presents parameters for county-region differences that have been adjusted by removing the statewide average difference. The first two columns present estimates in dollar terms, while the final two columns present estimates as a share of the mean pre-credibility-adjustment TM cost estimates for these counties.

Figure 5. Root Mean Squared Error of County TM Cost Estimates



Note: Lines depict the estimated RMSE of county TM cost estimates in several scenarios. Weighted RMSE reflects an average of estimated mean squared error across enrollment levels, weighted by the number of counties at each level. Gray bars report the percent of counties below the 1000-enrollee threshold that fall in each 100-enrollee interval.

pre-credibility-adjustment estimate for the affected counties. The extent of these differences varies widely across counties, with a standard deviation of $156 per enrollee per month, 15.0% of the mean pre-credibility-adjustment estimate. In short, these estimates suggest that the blending process engenders a bias that is negative on average, but varies widely from county to county.

Figure 5 combines our estimates of each of the terms in equation (1) to estimate the total error in the blended estimate, accounting for both sampling and non-sampling error. Specifically, the figure presents the root mean squared error (RMSE) expressed as a percent of the mean pre-credibility-adjustment estimate of TM costs for the affected counties.

For these counties, the actual RMSE of the blended estimate (solid dark blue line) is much larger than the RMSE accounting for sampling error alone (dashed dark blue line), which we estimate by setting the bias term in equation (1) to zero. The largest differences are for the lowest-enrollment counties, where the regional estimate receives the most weight and, thus, differences between counties and their regions are most important. Averaging across all counties with enrollment low enough to receive a credibility adjustment, the actual RMSE of the blended estimate is 6.5% of mean spending, compared to 3.5% accounting only for sampling error.

As noted earlier, after CMS produces the blended estimate $\tilde{Y}_c$, it applies a state-specific budget neutrality adjustment designed to ensure that the credibility adjustment process neither increases nor decreases its estimates of TM costs in each state on average. This budget neutrality adjustment may help to mitigate bias introduced by blending, so the final post-credibility-adjustment estimates may exhibit less error than the raw blended estimates examined in equation (1) and depicted by the solid dark blue line in Figure 5.

Directly estimating the error in CMS' estimates after the budget neutrality adjustment is not straightforward, so we instead consider a modified budget neutrality adjustment that is easier to model. Specifically, we subtract the statewide mean of the difference $Y_c^R - Y_c$ from the regional estimate Y_c^R prior to blending. Importantly, this modified budget neutrality adjustment likely outperforms CMS' actual approach. CMS' approach applies the same percentage adjustment to the blended estimate in all counties, so the adjustment will tend to be too large in counties with a low weight on the regional estimate and too small in counties with a high weight on the regional estimate. Our modified adjustment avoids this problem. Consequently, the error estimates we obtain for this modified budget neutrality adjustment should be viewed as a lower bound on the error that CMS achieves in practice using its actual budget neutrality adjustment.

To implement this approach, we first re-estimate the bias parameters Δ and τ^2 using the adjusted regional estimates.[20] Table 3 reports the resulting estimates. By construction, the mean county-region difference is now zero. But the *variation* in the county-region difference is only slightly smaller because most cross-county variation in these differences is within, not across, states.

We next plug these revised parameter estimates into equation (1) to obtain estimates of the RMSE under our modified budget neutrality adjustment, which are also reported in Figure 5.[21] The RMSE under our modified adjustment (light blue line) is lower than the RMSE of the raw blended estimate (solid dark blue line), but still much higher than the estimates that only account for sampling error (dashed dark blue line).

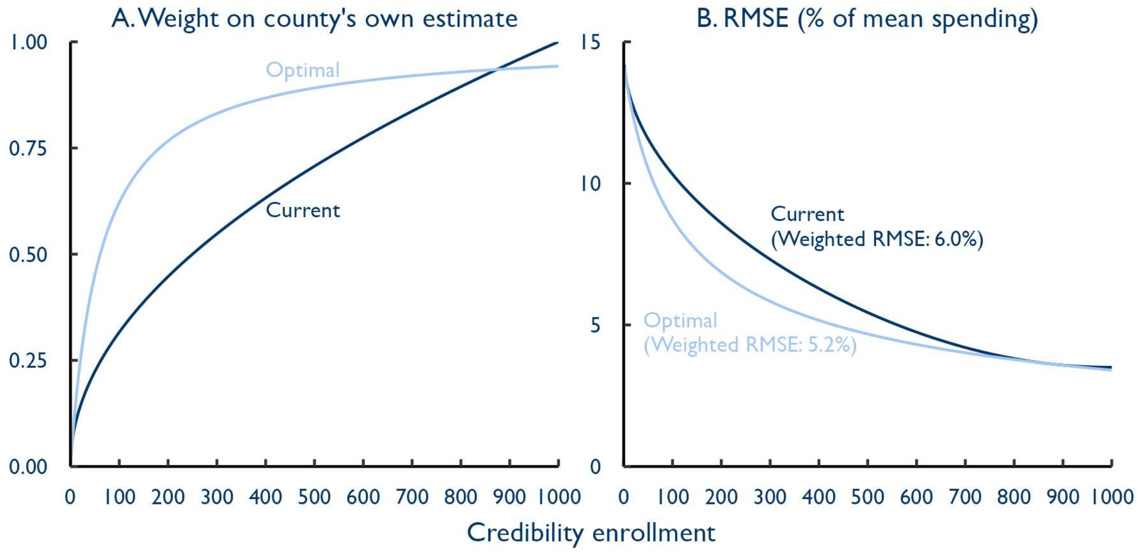
Estimation error under alternative weighting schemes

A final question is whether CMS could improve the credibility adjustment process. The results presented in Figure 6 suggest that this is possible; at many enrollment levels, the estimated RMSE without *any* credibility adjustment (green line) is lower than the post-adjustment RMSE (solid blue lines). That is, for many affected counties, the current adjustment *increases* error in expectation.

[20] In estimating τ, we increase the degrees of freedom used in calculating our estimate of $\text{Var}(Y_c^R - Y_c)$ to reflect the fact that we are now subtracting off state-specific means rather than a single pooled mean.

[21] Simply plugging these estimates back into equation (1) ignores the fact that the county-level estimates Y_c will be mechanically positively correlated with the adjusted regional estimate (because the regional estimate will be more likely to be adjusted upward when $Y_c - Y_c^R$ is positive). This will lead us to underestimate the mean squared error under this approach. In practice, this issue is likely relatively modest since most counties subject to a credibility adjustment are in states where a large number of counties are subject to an adjustment, and, as such, each individual county has little effect on the size of the adjustment made to the regional estimate.

Figure 6. Current versus Optimal Weighting Schemes



Note: Panel A plots the weight CMS currently assigns to the county's own estimate at each enrollment level and an estimate of the optimal weight derived in equation (4). Panel B plots estimates of RMSE of the resulting county TM cost estimates under the two weighting regimes. Both sets of estimates reflect the modified budget neutrality adjustment discussed in the text. Weighted RMSE reflects an average of estimated mean squared error across enrollment levels where each enrollment level is weighted by the number of counties with that enrollment.

One way CMS could change the credibility adjustment process is by changing the weight it assigns to the county's own estimate versus the regional estimate. We can use equation (1) to solve for the weight that minimizes the mean squared error of the resulting blended estimate. Differentiating equation (1) with respect to the weight $w(M_c)$ and solving the resulting first-order condition implies that the error-minimizing weight to place on a county's own estimate is:

$$w^*(M_c) = \frac{\mathbb{E}[\text{Var}(Y_c^R|\theta_c)|M_c] + \mathbb{E}[(\mu_c^R - \mu_c)^2|M_c]}{\mathbb{E}[\text{Var}(Y_c|\theta_c)|M_c] + \mathbb{E}[\text{Var}(Y_c^R|\theta_c)|M_c] + \mathbb{E}[(\mu_c^R - \mu_c)^2|M_c]}. \tag{4}$$

The optimal weight has the usual property that the weight assigned to each estimate depends on its mean squared error. The weight placed on the county estimate decreases with the mean squared error of the county's own estimate, which is simply the expected variance $\mathbb{E}[\text{Var}(Y_c|\theta_c)|M_c]$, and it increases with the mean squared error of the regional estimate, which is $\mathbb{E}[\text{Var}(Y_c^R|\theta_c)|M_c] + \mathbb{E}[(\mu_c^R - \mu_c)^2|M_c]$.

We plug the empirical estimates derived above into equation (4) to estimate the optimal weight and, in turn, the mean squared error under that weight. (In doing so, we use the estimates that reflect the modified budget neutrality adjustment described above.)[22] Panel A of Figure 6 shows that the optimal weight on the county's own estimate is higher than the current CMS weight, except for counties just below the 1,000-enrollee threshold at a credibility adjustment begins to apply. Panel B of Figure 6 compares the RMSE under the alternative weighting schemes. RMSE is lower under the optimal weighting scheme, particularly for counties with TM enrollment far below the 1,000-enrollee threshold. Averaging across all the counties with enrollment small enough to be subject to a credibility adjustment, RMSE under the optimal weight is 5.2% of mean spending versus 6.0% of mean spending under the current weight.

[22] The resulting calibrated version of equation (4) is $w^*(M_c) = 21953 \,/\, (1337269/M_c + 21953)$.

Discussion and Policy Implications

We find that sampling error in CMS' estimates of county TM costs is typically relatively modest. For the 2022 plan year, the median Medicare beneficiary lived in a county where the standard error of CMS' estimate was 0.6% of estimated TM costs, and even the 95th percentile beneficiary lived in a county where this standard error was 1.9% of TM costs. The amount of sampling error will likely grow only modestly in the coming years, with the median rising to 0.7% and the 95th percentile rising to 2.4% even in the lowest enrollment scenario we consider. This is partly because expected declines in TM's market share will be at least partly offset by demographically driven increases in total Medicare enrollment.

Our results do indicate that the credibility adjustments applied in counties with very low TM enrollment can introduce substantial non-sampling error. For counties subject to these adjustments, we estimate that the total root mean squared error of CMS' estimates—accounting for both sampling error and the non-sampling error introduced by the credibility adjustment process—was 6.5% of mean spending in 2022 versus 3.5% accounting for sampling error alone. However, only 1.2% of beneficiaries lived in counties subject to an adjustment in 2022, and only 2.2% of beneficiaries will live in such counties by 2033 in even the lowest-enrollment scenario we consider. Thus, non-sampling error introduced by the credibility adjustment process is of limited importance in the aggregate, although it is important in some areas.

Our overarching conclusion is that—at current TM enrollment levels and those likely to prevail for at least the next decade or so—CMS can construct fairly precise estimates of county TM costs, at least in areas where the large majority of Medicare beneficiaries live. Thus, concerns about the precision of these estimates offer, at best, a weak rationale for proposals that would partially or fully "decouple" MA payment from local TM costs, including proposals to base benchmarks on plan bids rather than TM costs (e.g., MedPAC 2023a; Ginsburg and Lieberman 2024) or to pay for some portion of plans' costs directly, as would occur under some reinsurance proposals.[23] This is particularly true of bidding proposals since MA plans would likely set their bids at least in part based on their own local claims experience and, thus, need to rely on similar or smaller samples. While changes to MA payment in this vein may indeed be advisable, they would need to be justified on other grounds (e.g., the limitations of the MA risk adjustment system).

There do appear to be opportunities to improve how CMS estimates county TM costs in counties with very low TM enrollment. Our results suggest that CMS could reduce the error of its blended estimates by determining how much weight is assigned to the regional estimate versus a county's own estimate using the "optimal" weight formula derived in this paper. CMS could likely also improve performance by replacing its current budget neutrality adjustment with the modified adjustment we model here; the current adjustment likely overcompensates in counties where the regional estimate gets little weight and undercompensates elsewhere. Future research could also explore whether there are ways to construct the regions used for blending in ways that create a closer match between counties and their regions.

[23] The same point applies to other policy changes that could ameliorate statistical problems posed by limited TM enrollment, such as changes that would seek to encourage enrollees to shift back into TM.

References

Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. 2024. "2024 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds." https://www.cms.gov/oact/tr/2024.

Centers for Medicare and Medicaid Services (CMS). 2020. "Advance Notice of Methodological Changes for Calendar Year (CY) 2022 for Medicare Advantage (MA) Capitation Rates and Part C and Part D Payment Policies – Part II." https://www.cms.gov/files/document/2022-advance-notice-part-ii.pdf.

———. 2021a. "Announcement of Calendar Year (CY) 2022 Medicare Advantage (MA) Capitation Rates and Part C and Part D Payment Policies." https://www.cms.gov/files/document/2022-announcement.pdf.

———. 2021b. "FFS Data 2019." https://www.cms.gov/files/zip/ffs-data-2019.zip.

Congressional Budget Office (CBO). 2024. "An Update to the Budget and Economic Outlook: 2024 to 2034." https://www.cbo.gov/publication/60039.

Freed, Meredith, Jeannie Fuglesten Biniek, Anthony Damico, and Tricia Neuman. 2024. "Medicare Advantage in 2024: Enrollment Update and Key Trends." KFF. https://www.kff.org/medicare/issue-brief/medicare-advantage-in-2024-enrollment-update-and-key-trends/.

Ginsburg, Paul, and Steven M. Lieberman. 2024. "Fixing Medicare Advantage With Competitive Bidding." USC Schaeffer Center. https://healthpolicy.usc.edu/research/fixing-medicare-advantage-competitive-bidding/.

Jacobson, Gretchen A., and David Blumenthal. 2022. "Medicare Advantage Enrollment Growth: Implications for the US Health Care System." *JAMA* 327 (24): 2393–94. https://doi.org/10.1001/jama.2022.8288.

Lieberman, Steven M., Paul B. Ginsburg, and Samuel Valdez. 2023. "Favorable Selection Ups The Ante On Medicare Advantage Payment Reform." https://doi.org/10.1377/forefront.20230606.520135.

McWilliams, J. Michael. 2022. "Don't Look Up? Medicare Advantage's Trajectory And The Future Of Medicare." *Health Affairs Forefront* (blog). March 24, 2022. https://www.healthaffairs.org/do/10.1377/forefront.20220323.773602/.

Medicare Payment Advisory Commission (MedPAC). 2023a. "Medicare and the Health Care Delivery System." https://www.medpac.gov/wp-content/uploads/2023/06/Jun23_MedPAC_Report_To_Congress_SEC.pdf.

———. 2023b. "Medicare Advantage Program Payment System." https://www.medpac.gov/wp-content/uploads/2022/10/MedPAC_Payment_Basics_23_MA_FINAL_SEC.pdf.

———. 2024. "Medicare Payment Policy." https://www.medpac.gov/document/march-2024-report-to-the-congress-medicare-payment-policy/.