CHAPTER 5

Improving the accuracy of Medicare Advantage payments by limiting the influence of outliers in CMS's risk-adjustment model

CHAPTER 5

Improving the accuracy of Medicare Advantage payments by limiting the influence of outliers in CMS's risk-adjustment model

Chapter summary

The Medicare program pays managed care plans that participate in Medicare Advantage (MA) a monthly capitated amount to provide Medicare-covered services to each of their enrollees. CMS adjusts the monthly capitated amounts for each enrollee using a risk score, which is a beneficiary-level index that indicates how costly an enrollee would have been expected to be in fee-for-service (FFS) Medicare relative to the national average FFS beneficiary. CMS generates the risk score for each enrollee using the CMS hierarchical condition category (HCC) model, a risk-adjustment model that uses demographic and medical information for FFS beneficiaries to predict the costliness of care.

The purpose of risk adjustment is to accurately predict costs not for a particular person, but rather on average for a group of people with the same attributes that affect health care costs. Risk-adjusted payments for some enrollees are less than their actual costs, while payments for others are higher than actual costs, but on average payments are accurate. The risk of financial loss provides an incentive for plans to manage their enrollees' conditions to keep their costs down. In addition, paying accurately for each condition on average reduces the incentives for plans to avoid enrolling beneficiaries with high-cost conditions.

In this chapter

The CMS–HCC risk-adjustment model has largely been successful in serving its general purpose. However, one ongoing concern is the inaccuracy introduced into the model by outliers—beneficiaries who have the largest differences between actual medical costs and the costs predicted by the model. Including these outlier costs in the CMS-HCC model biases the estimates of the model coefficients, which indicate the costs related to beneficiaries' demographic variables and medical conditions. The biased coefficients result in risk-adjusted payments that are too high for some enrollees and too low for others, which undermines the accuracy of payment to plans.

To address the inaccuracy introduced in the CMS–HCC model by outliers, we evaluated a modification that incorporates the principles of reinsurance and repayment to limit the influence of outliers in the estimation of the model's coefficients. The modification, developed by Tom McGuire, Sonja Schillo, and Richard van Kleef, uses financial transfers to redistribute payments from plans whose enrollees incurred costs substantially below the model's prediction to plans whose enrollees incurred costs substantially above predicted costs. However, these redistributions are not administratively feasible in Medicare Advantage. Therefore, we used the fundamental ideas from McGuire, Schillo, and van Kleef with minor changes designed to minimize the impact on the process used to risk adjust payments to MA plans. Our method would address outliers in the risk-adjustment model using simulation principles but would not require any change in the flow of funds from CMS to MA plans (i.e., it would not require payment withholds, subsequent reconciliations, or changes to the payment process).

In our method, we divided our analytic sample into estimation and evaluation subsamples. Using the estimation sample, we calculated the difference between the actual costs and the costs predicted by the standard CMS–HCC model for each FFS beneficiary in our analytic file. For beneficiaries with the largest underpredictions (predicted costs less than actual costs), we applied a loss limit by reducing their actual costs (similar to a reinsurance payment) such that the total reduction in costs equaled 2 percent of all costs. For beneficiaries with the largest overpredictions (predicted costs greater than actual costs), we applied a gain limit by increasing the actual costs (similar to requiring a repayment) such that the total increase in costs equaled 2 percent of all costs. The limits offset one another so that the adjustment to the cost data is revenue neutral. We then used the adjusted cost data (with loss and gain limits applied) to re-estimate the CMS–HCC model, thereby limiting the influence of outliers on the resulting coefficients with no impact on the flow of funds from CMS to MA plans.

We used the evaluation sample to evaluate the effect of this method of limiting overpredictions and underpredictions. We used two measures of overall fit, the R^2 and the Cumming's prediction measure, both of which estimate how well predicted costs reflect actual costs. We found that this modification to the standard CMS–HCC model improved the R^2 from 0.13 to 0.19. For context, since 2007 (the first year the CMS–HCC model was fully implemented), all model changes to improve accuracy have increased the R^2 from about 0.11 to 0.13. For the Cumming's prediction measure, the improvement is smaller, rising from 0.13 to 0.16. In addition, we assessed how well the modified model predicts costs for groups of beneficiaries using the predictive ratio, which is the aggregate costs for the group predicted by the risk-adjustment model divided by the aggregate actual costs for the group. We considered groups of beneficiaries for which the standard CMS–HCC model performs less well (those with very low and very high actual costs and those with very large underpredictions and overpredictions) and found improvements in model performance.

Improving the accuracy of MA risk adjustment is a goal for the Commission. This approach would help accomplish that goal without any additional burden on plans or beneficiaries to provide additional data. Further, CMS could continue to use a risk-adjustment model that is familiar, straightforward, and easy to understand. But substantial issues would remain for MA risk adjustment, such as the financial benefit to plans for coding conditions more intensively relative to FFS clinicians' coding and the payment inaccuracies among beneficiaries who are not among the largest overpredictions and underpredictions addressed in this analysis. In addition, more work is needed to understand how this approach can integrate with other improvements to risk adjustment for MA plans. The Commission intends to address these issues in future work. ■

Background

Medicare pays managed care plans that participate in the Medicare Advantage (MA) program a monthly capitated amount for each enrollee to provide Medicare-covered services. Each capitated payment has two parts: a base rate and a risk score. CMS determines a plan's base rate using the plan's bid and county benchmarks for the plan's service area. CMS standardizes the base rates using the health status of the national average beneficiary in fee-for-service (FFS) Medicare.[1] CMS then uses a risk score to adjust the standardized base rate for an MA plan up or down for each enrollee, depending on the enrollee's health status relative to the national average. The risk scores are beneficiary-level indexes that indicate the expected Medicare costs for an enrollee relative to the national average FFS beneficiary. How well Medicare's payments to MA plans match their enrollees' costliness depends in large part on how well the risk scores predict the expected costs for the plans' enrollees.

Medicare spending varies widely among beneficiaries. Some of this variation is predictable because it depends on beneficiary characteristics that can be observed, such as age, chronic medical conditions, or historical health care use. The rest of the variation is generally not predictable from information that CMS has available because the variation is due to random medical events, such as a heart attack or hip fracture. The base rates reflect the costs of random events that are part of the MA payments. Risk-adjustment models strive to address predictable spending variation because otherwise MA plans could use beneficiaries' observable characteristics to their advantage through favorable selection—avoiding beneficiaries with certain (unprofitable) attributes and attracting those with favorable (profitable) attributes.

The general purpose of risk adjustment is to accurately predict costs not for a particular person but on average for a group of people with the same attributes that affect health care costs (Newhouse et al. 2012). For enrollees who have the same risk score, payments will be below actual costs for some (that is, the risk model will underpredict costs) and above actual costs for others (that is, the risk model will overpredict costs) but will be accurate on average. This result is a feature of all models that use patients' conditions to predict costs. While the risk of financial losses provides an incentive for plans to manage their enrollees' conditions to keep their costs down, the risk-adjustment model should avoid systemic underpredictions or overpredictions.

Risk adjusting MA payments: The CMS–HCC model

Over the years, CMS has used a variety of methods for determining MA enrollees' risk scores. Currently, CMS uses the CMS hierarchical condition category (CMS–HCC) risk-adjustment model, which uses enrollees' demographic characteristics and medical conditions (such as diabetes and stroke) to predict their costliness. CMS draws data for demographic variables—which include age, sex, level of Medicaid benefits (if any), institutional status, eligibility based on disability, and eligibility based on age but originally eligible because of disability—from the year in which beneficiaries' costs are to be predicted (the prediction year). CMS bases each beneficiary's medical conditions (such as diabetes and stroke) on diagnoses recorded on physician, hospital outpatient, and hospital inpatient claims from the year before the prediction year (base year). The CMS–HCC model is prospective, meaning it uses conditions from a base year to predict beneficiary costs in the next year (the prediction year).

CMS groups the diagnoses into broader disease categories called HCCs. In the CMS–HCC model, CMS has aligned some conditions with more than one HCC, which differ by severity of the condition, and CMS has arrayed them in a hierarchy. For example, the CMS–HCC model has three HCCs for diabetes: without complications, with chronic complications, and with acute complications. The "hierarchical" aspect of HCCs means that if a beneficiary's diagnoses map into more than one HCC in a condition hierarchy, CMS applies only the HCC that has the largest effect on the beneficiary's risk score—the highest-severity HCC.

Each demographic and HCC component in the risk-adjustment model has a coefficient that represents the expected medical costs associated with that component. CMS estimates these coefficients using FFS Medicare claims data such that all Medicare spending in a year is distributed among the model components. CMS sums the coefficients from the demographic and HCC components that apply to a beneficiary to create the beneficiary's predicted cost.

CMS calculates a risk score for the beneficiary by dividing the beneficiary's predicted cost by the cost of the national average FFS Medicare beneficiary. Hence, the risk score indicates the percentage difference between the beneficiary's expected cost and the cost of the national average FFS beneficiary. For example, if a beneficiary has a risk score of 1.65, the beneficiary's expected cost is 65 percent higher than the national average cost.

Optimizing risk adjustment for payment accuracy

CMS regularly updates the CMS–HCC model with more recent data to ensure that the risk scores reflect recent treatment costs. In addition, since the full implementation of the CMS–HCC risk-adjustment model in 2007, CMS has made the following modifications to improve how well the model predicts health care costs:

- ***Revised the mapping of diagnosis codes to HCCs.*** A team of clinicians developed a mapping of all International Classification of Diseases, Ninth Revision, Clinical Modification (ICD–9–CM) diagnosis codes to HCCs and, later, a mapping of all ICD–Tenth Revision (10)–CM diagnosis codes to HCCs. Over time, CMS has revised this mapping to group diagnosis codes into more similar groups based on treatment costs or diagnosing patterns.
- ***Added and deleted HCCs.*** In developing the model, CMS determined which HCCs influence overall health care costs for a beneficiary in the prediction year by identifying those that improve the overall predictive power of the model or that improve payment accuracy for certain groups when included in the model. CMS includes in the model only those HCCs that meet a threshold of influence and other criteria (such as diagnostic specificity). Over time, CMS has made improvements to the model by adding and deleting individual HCCs.
- ***Added a count of HCCs.*** The CMS–HCC model has always included "interactive variables" that are designed to address the higher costs that sometimes occur when a beneficiary has multiple conditions or when beneficiaries who are disabled have certain conditions. For example, the average cost of treating beneficiaries with diabetes *and* congestive heart failure is higher than the sum of the average cost of treating beneficiaries with diabetes only and the average cost of treating beneficiaries with congestive heart failure only. In addition to the existing interactive variables, CMS is phasing in a set of HCC count variables to address the higher costs that occur for beneficiaries with four or more HCCs.
- ***Stratified populations.*** Initially, CMS used distinct versions of the CMS–HCC model for new enrollees (who do not have a full calendar year of diagnostic data), beneficiaries with end-stage renal disease, and all other beneficiaries. For all other beneficiaries, CMS used models that calculated separate risk scores for beneficiaries residing in the community and for those residing in an institution (on a monthly basis) and used a set of variables within the model to account for the higher health care costs of beneficiaries who are disabled and those who are eligible for Medicaid benefits (dually eligible beneficiaries). In 2017, CMS significantly improved the model's accuracy by stratifying community-residing beneficiaries based on eligibility for Medicaid benefits (full, partial, or no benefits) and Medicare eligibility status (beneficiaries 65 or older are eligible based on age and younger beneficiaries are eligible based on disability).

In general, the CMS–HCC model succeeds at avoiding systemic underpayments and overpayments for many populations (Medicare Payment Advisory Commission 2020). The changes by CMS since full implementation of the CMS–HCC model in 2007 have improved how well the model predicts costs for specific groups. CMS has not yet tried to improve how well the model performs for the population as a whole, which is reflected in the small increase in the model's R^2 (a statistical measure of how much of the variation in beneficiaries' costliness is explained by the risk-adjustment model) from 0.11 to 0.13 as a result of these improvements. In this chapter, we seek to improve the model's performance for the whole population by reducing the influence of outliers—beneficiaries with the largest prediction errors in the current risk-adjustment model—on the model's coefficients, thereby improving payment accuracy.

Outlier costs in CMS's risk-adjustment model undermine payment accuracy

CMS uses a regression to estimate the size of each demographic and HCC coefficient in the CMS–HCC model. The model estimation essentially allocates each FFS beneficiary's annual Medicare costs to the model's demographic and HCC components such that all costs are accounted for by the model coefficients applied to a given beneficiary. Because the regression is run using all FFS beneficiaries, each coefficient represents the average annual Medicare spending (across all FFS beneficiaries) associated with the demographic characteristic or HCC.

A small share of beneficiaries have annual Medicare costs that are very high or very low. Including these outlier beneficiaries in the risk model estimation introduces bias in the coefficients and generates payment inaccuracy. Consider a simplified hypothetical example in which a beneficiary who is 75 and has four HCCs and actual medical costs of $1.5 million is added to the model estimation population. In this case, the age coefficient and each of the four HCC coefficients are associated with about $300,000 in costs ($1.5 million / 5 coefficients = $300,000 on average). If one of these HCCs has 10,000 other beneficiaries with an average cost of $3,000 associated with the HCC, adding the outlier beneficiary to the estimation increases the average HCC cost to $3,030 (calculation: [(10,000 × $3,000 + 1 × $300,000) / 10,001] = $3,030).[2]

In addition to introducing bias to the coefficients applicable to the outlier beneficiary, the coefficients for other HCCs are also biased by adding the outlier beneficiary to the model estimation population. CMS first calculates HCC coefficients in dollar terms, then divides them by the average annual Medicare costs among FFS beneficiaries (so that that average risk score is 1.0). Adding the outlier beneficiary increases that average annual Medicare cost so that other HCC coefficients are lower. For example, an HCC with average costs of $2,000 has a coefficient of 0.200 when the average annual Medicare cost is $10,000 but has a coefficient of 0.199 if the per capita average annual Medicare costs increase to $10,001.

For simplicity, we used a beneficiary with very high annual medical costs in the hypothetical example above. However, the beneficiaries who cause the most bias in coefficients are those with the largest differences in predicted and actual costs. (Model coefficients would not be biased by a beneficiary with very high costs if those costs were perfectly accounted for by the model's variables.) We note that both large overpredictions (predicted costs much larger than actual costs) and underpredictions (predicted costs much smaller than actual costs) can bias coefficients.

Finally, as we explained earlier, risk adjustment seeks to accurately predict average costs for a group of beneficiaries with the same attributes. If the risk model's coefficients are biased (some produce overpayments and others produce underpayments), a plan could experience financial gains or losses depending on whether their enrollees disproportionately have coefficients that overpay or underpay. The potential gains or losses give plans incentive to attract or avoid beneficiaries who have particular conditions, based on whether the condition has a coefficient that is too high or too low.

Using principles of reinsurance and repayment to limit the effect of outlier prediction errors in the risk-adjustment model

Many insurance markets use a system of reinsurance and repayments to address beneficiaries with outlier costs. Reinsurance provides additional payments for plan enrollees with medical costs that are much greater than premium payments, and repayments recoup payments from plans for enrollees whose medical costs are much less than premium payments to plans. In these markets, reinsurance and repayments often operate as a system of financial transfers that occur after an initial set of premium payments to plans. In MA, however, medical cost data are not available to serve as a basis for determining reinsurance and repayment amounts.

McGuire, Schillo, and van Kleef developed a method of reinsurance and repayment and described how it would improve risk adjustment in several health insurance markets (McGuire et al. 2020). However, the method used by McGuire, Schillo, and van Kleef would require explicit redistributions of payments among MA plans, which is not administratively feasible. Instead, we utilized a modified version of their method. This method incorporates the principles of reinsurance and repayment and focuses on the largest prediction errors (both overpredictions and underpredictions), in which the difference between the annual cost for a beneficiary predicted by the model (through a risk score) and the actual annual cost for that beneficiary is large. But rather than using actual financial transfers to redistribute payments among plans through a system of reinsurance and repayments, the method redistributes a share of annual beneficiary costs in the FFS data used to estimate the risk-adjustment model coefficients and does not require any change to the flow of funds from CMS to MA plans (i.e., no payment withholds, subsequent reconciliations, or changes to the payment process).

The redistribution of costs in the FFS data targets the most extreme prediction errors, affecting a small fraction of beneficiaries. However, these extreme prediction errors can distort the model coefficients and reduce the accuracy of all beneficiaries' risk scores. For beneficiaries with the largest underpredictions (predicted costs lower than actual costs), we applied a loss limit such that if hypothetical reinsurance payments were provided to cover all losses above the loss limit, the total amount of reinsurance payments would equal 2 percent of all costs. Similarly, for beneficiaries with the largest overpredictions (predicted costs higher than actual costs), we applied a gain limit such that if hypothetical repayments were required to recoup all gains above the gain limit, the total amount of repayments would equal 2 percent of all costs. The size of the redistribution of costs could be smaller or larger, but the redistribution should net to zero so that the modification of risk model estimation is revenue neutral.

A benefit of this approach to addressing large prediction errors is that it would improve the performance of the CMS–HCC model without additional burden on plans and beneficiaries to provide additional data. In addition, CMS would continue to use a risk-adjustment model that is familiar, straightforward, and easy to understand.

To apply the loss and gain limits, we adjusted the actual cost data for the affected beneficiaries so that the beneficiary's prediction error does not exceed the loss or gain limit. After we applied the cost redistribution to beneficiaries' cost data, we re-estimated the coefficients for all the model's variables using the redistributed cost data, which optimizes all the model's coefficients and improves the accuracy of risk scores for all beneficiaries.

We used the following steps to implement this method:

1. We divided our analytic sample into two subsamples: The estimation sample and the evaluation sample.
2. Using the estimation sample, we estimated coefficients for all variables in the standard CMS–HCC model using actual (nonredistributed) cost data.
3. We predicted costs for each beneficiary on the estimation sample using the coefficients from (2) and then calculated prediction errors by subtracting actual costs from predicted costs for each beneficiary.
4. We applied a loss limit to the beneficiaries with the largest underpredictions—largest differences in actual costs minus predicted costs—by reducing actual costs so that the total adjustment equals 2 percent of all costs.
5. We applied a gain limit to the beneficiaries with the largest overpredictions—largest differences in predicted costs minus actual costs—by increasing actual costs so that the total adjustment equals 2 percent of all costs.
6. We re-estimated coefficients for all the model's variables using the redistributed cost data (with loss and gain limits applied), producing the optimized coefficients.
7. Using the evaluation sample, we evaluated how well the model with optimized coefficients performs in terms of predictive accuracy.

The text box provides a detailed description of our method.

Method for analyzing the effects of limiting outliers in the CMS hierarchical condition category model

In our analysis evaluating the effects of limiting outliers in the CMS hierarchical condition category (CMS–HCC) model estimation, we used a version of the CMS–HCC model that CMS used to risk adjust Medicare Advantage (MA) payments in 2019 for beneficiaries who were age 65 or older and not eligible for both Medicare and Medicaid. CMS named this version of the CMS–HCC model V23.[3] Use of V23 required us to use an analytic file that included Medicare beneficiaries who met these criteria:

- enrolled in both Part A and Part B of fee-for-service (FFS) Medicare during all 12 months of 2018,
- enrolled in FFS Medicare for at least one month in 2019,
- eligible for Medicare based on age (age 65 or older),
- not indicated as being in a long-term care institution, and
- did not receive Medicaid benefits.

We used our analytic file of 20.4 million beneficiaries in a seven-step method to estimate and evaluate the effects of limiting outliers. In step 1, we divided the analytic file into two files—an estimation sample and an evaluation sample—that had approximately the same number of records (about 10.2 million). We used the estimation sample for steps 2-6 and the evaluation sample for step 7.

In step 2, we used the estimation sample in a weighted least squares regression to estimate a standard version of V23. The explanatory variables in model V23 include 14 age/sex categories (7 age categories for men and 7 age categories for women); 83 HCCs; 6 disease interaction terms; and 2 indicators for whether a beneficiary was originally eligible for Medicare based on disability (one for men, one for women).

We used the coefficients from the estimated V23 model to produce predicted costs for each beneficiary in the estimation file. We calculated the difference between each beneficiary's actual costs and their predicted costs—the prediction error—which indicates how much a plan would gain or lose financially on that beneficiary (step 3). We sorted beneficiaries with an underprediction from largest to smallest, and through an iterative process we identified a *loss limit* of $106,512. We reduced the actual costs of beneficiaries who had underpredictions greater than the loss limit (0.4 percent of beneficiaries in the estimation file) so that the sum of the cost reductions equaled 2 percent of the total costs among all beneficiaries in our estimation file (step 4).[4]

We sorted beneficiaries with an overprediction from largest to smallest overprediction and used an iterative process to identify a gain limit of $25,268. We increased the actual costs of beneficiaries who had overpredictions greater than the gain limit (1.8 percent of beneficiaries in the estimation file) so that the sum of the cost increases equaled 2 percent of the total costs among all beneficiaries in our estimation file (step 5).[5] We made no adjustments to actual costs for beneficiaries not affected by the loss or gain limits.

We used the adjusted costs in a new regression to re-estimate the V23 model. We performed an iterative process in which we calculated new loss and gain limits with each iteration, calculated adjusted costs based on the new limits, and then re-estimated the V23 model based on the new adjusted costs. We continued the iterative process until the change in the loss and gain limits was less than $1 from one iteration to the next (step 6).[6]

Finally, we used the evaluation sample to evaluate how well the two models that we estimated—the standard CMS–HCC model and the re-estimated version with the adjusted costs—perform in terms of predicting beneficiaries' costs (step 7). ■

TABLE 5–1

Risk-adjustment model prediction improves when using a modified model that limits the influence of outlier predictions

Statistical measure	Result from standard model	Result from modified model
R^2	0.13	0.19
Cumming's prediction measure	0.13	0.16

Note: The modified model is designed to mitigate the effects of unusually large underpredictions and overpredictions on the model's predictive power. A higher R^2 and Cumming's prediction measure indicate better predictive power.

Source: MedPAC analysis of the version of the CMS hierarchical condition category (HCC) model that CMS used to risk adjust Medicare Advantage payments in 2019 and analysis of that version of the CMS–HCC model with a modification to limit the effects of outliers. Data used in this analysis include standard analytic claims files from 2018, the Common Medicare Enrollment file, and the Medicare risk-adjustment file for 2019.

Limiting the effect of outlier predictions would substantially improve the model's predictive power

The purpose of this analysis was to evaluate how well adjusting the costs during model estimation for beneficiaries who have the largest underpredictions and overpredictions would improve the predictive power of the CMS–HCC model. To evaluate how well predicted costs aligned with actual costs across the beneficiaries in our analytic file, we used two measures of overall model fit: the R^2 statistic and the Cumming's prediction measure (CPM), which is a linear version of the R^2.[7] To calculate the R^2 for both the standard model and the modified model, we simply used the R^2 statistics produced by the regressions that we used to estimate the coefficients for these two models.[8]

We found that the version of the modified V23 model, which limits the influence of outliers, performed better than the standard version of V23. The R^2 we calculated from our evaluation file increased from 0.13 under the standard model to 0.19 under the modified version, a 43 percent increase (Table 5-1). This result indicates that the modified version accounts for 43 percent more variation in beneficiaries' actual costs compared with the variation accounted for by the standard version, a major improvement in model accuracy.[9] We also found that the CPM increased by a smaller amount—from 0.13 to 0.16—from the standard version to the modified version (Table 5-1).[10]

The improvement in predictive power as measured by the R^2 is larger than all of CMS's prior model enhancements since 2007 combined, which collectively increased the share of cost variation explained by the model from 0.11 to 0.13 (Centers for Medicare & Medicaid Services 2021). The improvement in the R^2 observed with the modified V23 version is also larger than the increase resulting from other suggested changes to the CMS–HCC model. For example, Frogner and colleagues evaluated the effects of adding the number of chronic conditions for each beneficiary and using two years of diagnosis data to determine beneficiaries' HCCs rather than a single year of data, which CMS currently uses (Frogner et al. 2011). By incorporating the number of chronic conditions for each beneficiary and using two years of diagnosis data in the model they evaluated, they increased the R^2 of the model from 0.101 to 0.104.[11]

By limiting the influence of outliers (reducing the largest prediction errors, which account for much of the aggregate prediction errors), our method produces greater improvement in R^2 relative to the prior improvements to the CMS–HCC model, which focused on subsets of the Medicare population that may or may not have large prediction errors.

Modified CMS–HCC model improved cost prediction for the costliest beneficiaries and for those with large prediction errors

The primary purpose of limiting the influence of outliers during model estimation is to improve payment accuracy for the beneficiaries whose costs reflect the largest prediction errors, so it is vital that the modified model perform better than the standard model for the largest errors. To evaluate how well the models predict costs for these outliers, we calculated predictive ratios (PRs) for beneficiaries who have the lowest and highest spending and the largest prediction errors (both underpredictions and overpredictions) under the standard model and the modified model. PRs are the ratio of costs for a group of beneficiaries predicted by a risk-adjustment model to the actual costs for that group. PRs less than 1.0 indicate that the model underpredicts the costs for that group, and PRs greater than 1.0 indicate that the model overpredicts costs. The closer the PR is to 1.0 for a group of beneficiaries, the better the model has predicted the costs for the group. For example, both a PR increase from 0.8 to 0.9 and a PR decrease from 1.2 to 1.1 represent improvements in the model's prediction.

We stratified beneficiaries' actual Medicare costs used to estimate the standard CMS–HCC model into percentiles. We found that, relative to the standard model, the modified model produced small improvements in the PRs for most of the spending categories. For example, for the beneficiaries who have spending between the 40th percentile and 60th percentile, the PR decreased by 8 percent from 2.99 under the standard model to 2.76 under the modified model (Table 5-2, p. 152). However, the improvement for the beneficiaries with the highest spending (the 99th percentile) was greater: The PR increased by 28 percent, from 0.14 to 0.18. We also found improvements in PRs under the modified model for beneficiaries who had the largest underpredictions and overpredictions under the standard model. For the beneficiaries with the 1 percent largest underpredictions, the PR improved by 20 percent, from 0.13 to 0.16. For the beneficiaries with the 1 percent largest overpredictions, the PR improved from 6.4 to 2.0, a decrease of 68 percent. The strong improvement in the PRs for the beneficiaries who have the highest spending, the largest underpredictions, and the largest overpredictions indicates that the modified version substantially improves cost predictions for outliers, which leads to better overall model performance.

Studies have used PRs to show that the standard CMS–HCC model predicts costs very accurately for groups of beneficiaries defined by their medical conditions (Centers for Medicare & Medicaid Services 2018, Centers for Medicare & Medicaid Services 2016, Medicare Payment Advisory Commission 2020, Medicare Payment Advisory Commission 2014, Medicare Payment Advisory Commission 2012, Pope et al. 2004). To allay concerns that the modified model might not perform as well as the standard model in terms of predicting costs for specific medical conditions, we evaluated how well both models perform in predicting total costs for beneficiaries who have any of these common conditions: cancer, acute myocardial infarction, diabetes, congestive heart failure, chronic obstructive pulmonary disease, or stroke. For each of these conditions, the PR is 1.0 under both the standard model and the modified model. These results indicate that both versions of the model pay accurately, on average, for beneficiaries who have these conditions (Table 5-2, p. 152).

The modified CMS–HCC model explained a greater amount of cost variation for 15 common medical conditions

We have shown that limiting outlier prediction errors reduces the extent of large overpredictions and underpredictions under the CMS–HCC model and improves model performance overall. Given these results, we are certain that optimized coefficients reduce aggregate prediction errors across all HCCs; however, it is possible that the optimized coefficients increase prediction errors for a minor share of beneficiaries. In this situation, there is a theoretical concern that the optimized coefficient for an individual HCC produces larger aggregate prediction errors than the standard model coefficient, thereby allowing more opportunities for plans to attract favorable risks and avoid unfavorable risks.

We investigated how the method for limiting outlier prediction errors affects the amount of cost variation explained in 15 common HCCs. We found that in all 15 HCCs, the amount of cost variation explained is higher (the prediction errors are lower) under the modified model relative to the standard model. Using CPM as the measure, we found that the increase in variation

TABLE 5–2

Limiting the influence of outlier predictions would improve model performance for the costliest beneficiaries and for those with large prediction errors

Beneficiary category	Predictive ratio from standard model	Predictive ratio from modified model
Percentile of payment-year costs		
0 to 10th percentile	32.05	24.69
10th to 20th percentile	10.80	9.57
20th to 40th percentile	5.55	5.03
40th to 60th percentile	2.99	2.76
60th to 80th percentile	1.54	1.46
80th to 95th percentile	0.59	0.59
95th to 99th percentile	0.27	0.28
99th percentile or higher	0.14	0.18
Prediction error		
1% largest underpredictions	0.13	0.16
1% largest overpredictions	6.4	2.0
Conditions		
Cancer	1.00	1.00
AMI	1.00	1.00
Diabetes	1.00	1.00
CHF	1.00	1.00
COPD	1.00	1.00
Stroke	1.00	1.00

Note: AMI (acute myocardial infarction), CHF (congestive heart failure), COPD (chronic obstructive pulmonary disease). “Payment-year costs” are beneficiaries’ costs from the year in which Medicare Advantage payments would be determined. The table shows predictive ratios (PRs) for total Medicare costs for several categories of beneficiaries. PRs closer to 1.0 indicate greater accuracy. PRs below 1.0 indicate underpredictions on average for the category, and PRs greater than 1.0 indicate overpredictions on average for the category.

Source: MedPAC analysis of the version of the CMS hierarchical condition category (HCC) model that CMS used to risk adjust Medicare Advantage payments in 2019 and analysis of that version of the CMS–HCC model with modifications to reduce the effects of outliers. Data used in this analysis include standard analytic claims files from 2018, the Common Medicare Enrollment file, and the Medicare risk-adjustment file for 2019.

explained ranged from about 11 percent for HCC 12 (breast, prostate, and other cancers) to about 35 percent for HCC 86 (acute myocardial infarction) (Table 5-3). The key point, however, is that the amount of cost variation explained increased under the modified model in all 15 HCCs.

Effects of using different gain and loss limits

We also evaluated how our results would differ if we used different gain limits and loss limits that would result in different percentages of total costs being redistributed during model estimation. Specifically, we evaluated the effects of gain and loss limits that would result in 1 percent and, separately, 3 percent of costs being redistributed during model estimation.

Effects of redistributing 1 percent of total spending during model estimation

We found that a system that redistributes 1 percent of total spending during model estimation would require a loss limit of $147,617 and a gain limit of $30,635. Under

TABLE 5–3

Limiting the influence of outlier predictions would improve how well the CMS–HCC model predicts costs for 15 common HCCs

HCC	Number of beneficiaries (thousands)	CPM using standard model	CPM using modified model	Percent change in CPM
Lymphoma and other cancers	175	0.094	0.115	22%
Breast, prostate, and other cancers	781	0.097	0.108	11
Diabetes with chronic conditions	1,557	0.129	0.160	24
Morbid obesity	519	0.135	0.167	23
Inflammatory bowel disease	111	0.113	0.150	33
Rheumatoid arthritis	748	0.100	0.126	26
Major depressive, bipolar, and paranoid disorders	704	0.135	0.165	22
Parkinson's and Huntington's disease	153	0.091	0.114	25
CHF	1,213	0.100	0.132	32
AMI	186	0.122	0.165	35
Ischemic or unspecified stroke	289	0.109	0.142	30
Vascular disease	1,530	0.107	0.136	27
COPD	1,182	0.120	0.155	29
Exudative macular degeneration	232	0.053	0.070	32
Chronic kidney disease, stage 4	91	0.080	0.098	22

Note: HCC (hierarchical condition category), CPM (Cumming's prediction measure), CHF (congestive heart failure), AMI (acute myocardial infarction), COPD (chronic obstructive pulmonary disease). A higher CPM indicates better predictive power. Under the modified model, we reduced the cost for beneficiaries who had the largest underpredictions of costs and increased the cost for beneficiaries who had the largest overpredictions of costs.

Source: MedPAC analysis of the version of the CMS–HCC model that CMS used to risk adjust Medicare Advantage payments in 2019 and analysis of that version of the CMS–HCC model with modifications to reduce the effects of outliers. Data used in this analysis include standard analytic claims files from 2018, the Common Medicare Enrollment file, and the Medicare risk-adjustment file for 2019.

these parameters, the predictive power of the CMS–HCC model improves relative to the standard model: The R^2 improved by 24 percent—from 0.13 to 0.16—and the CPM improved by 10 percent—from 0.13 to 0.15. As expected, these improvements were smaller than under the method of setting the loss and gain limits at 2 percent of total spending, under which the R^2 improved by 43 percent and the CPM improved by 21 percent (Table 5-4, p. 154).

Effects of redistributing 3 percent of total spending during model estimation

We found that a system that redistributes 3 percent of total spending during model estimation would require a loss limit of $86,367 and a gain limit of $21,836. Under these parameters, the predictive power of the CMS–HCC model improves relative to the standard model: The R^2 improved by 61 percent—from 0.13 to 0.21—and the CPM improved by 32 percent—from 0.13 to 0.17. These increases were larger than when reinsurance and repayment amounts were 2 percent of total spending (Table 5-4, p. 154).

If CMS were to implement a system that redistributes costs from the largest underpredictions to the largest overpredictions during model estimation, the decision about the share of total costs to be redistributed should balance two effects. On the one hand, the model would have greater predictive accuracy, resulting in smaller prediction errors for the largest underpredictions and

TABLE 5–4

Predictive power of modified CMS–HCC model improves as size of reinsurance and repayments increases, but accuracy of model coefficients may be adversely affected

		Sensitivity analysis	Original analysis	Sensitivity analysis
Statistical measure	**Standard model**	**Model with 1% reinsurance/ repayment**	**Model with 2% reinsurance/ repayment**	**Model with 3% reinsurance/ repayment**
R^2	0.13	0.16	0.19	0.21
Cumming's prediction measure	0.13	0.15	0.16	0.17

Note: HCC (hierarchical condition category). The modified models are designed to mitigate the effects of unusually large underpredictions and overpredictions on the model's predictive power. A higher R^2 and Cumming's prediction measure indicate better predictive power.

Source: MedPAC analysis of the version of the CMS–HCC model that CMS used to risk adjust Medicare Advantage payments in 2019 and analysis of that version of the CMS–HCC model with a modification to limit the effects of outliers. Data used in this analysis include standard analytic claims files from 2018, the Common Medicare Enrollment file, and the Medicare risk-adjustment file for 2019.

overpredictions. On the other hand, redistributing costs during model estimation affects the coefficients on HCCs. As more costs are redistributed, the possibility increases that HCC coefficients would no longer accurately reflect the cost of treating the conditions represented by some HCCs.

Discussion and future work

One of the benefits of MA's capitated payments is that they provide incentives for plans to efficiently manage the care of their enrollees. Some may argue that applying a system of reinsurance and repayment would counteract the beneficial incentives created by capitated payments because it increases payments to plans in instances in which the cost of care exceeds the capitated payment. However, under the method we presented for addressing outliers—which uses principles of reinsurance and repayment without explicit payment adjustments—adjustments to the model coefficients generate significant improvements in payment accuracy overall, so we do not believe the incentive to manage care efficiently would be diminished. The number of enrollees whose costs would be adjusted is small because the adjustment to actual costs would occur only when model underpredictions or overpredictions are very large. In our analysis, we adjusted actual costs (either increased or decreased) for only 2.2 percent of the beneficiaries in our analytic file, yet the improvement in share of cost variation explained by the model increased from 0.13 to 0.19 in our main analysis. We note that this improvement in accuracy is several times larger than all prior model improvements combined, which collectively improved the share of cost variation explained by the model from about 0.11 to 0.13.

We restricted this analysis to beneficiaries who are ages 65 and older and did not have any Medicaid benefits (that is, we excluded beneficiaries dually eligible for Medicare and Medicaid) during the year of our analysis (2019). This population is one of seven population stratifications for which CMS has developed distinct versions of the CMS–HCC model. We used the age 65 and older population for this analysis because it is by far the largest of the seven population stratifications. If the approach we evaluated is to be effectively implemented in the risk adjustment for MA plans, the effect of this approach on the versions of the CMS–HCC model specific to the other six population stratifications should be evaluated.

In addition, we used HCCs in this analysis that CMS has defined using ICD–9–CM diagnosis codes. However, more precise ICD–10–CM diagnosis codes have been used extensively throughout the health care sector for a number of years. We encourage CMS to use ICD–10–CM codes to recalibrate the CMS–HCC model. The use of the ICD–10–CM codes will likely have some effect on the CMS–HCC coefficients, and the effects of the approach we evaluated in this study should be evaluated in the context of that recalibrated model.

Though this approach, which is only one way to address outliers, would improve model performance, substantial issues remain for MA risk adjustment, such as the financial benefit to plans of simply coding medical conditions more intensively relative to FFS clinicians' coding and the payment inaccuracies among beneficiaries who are not among the largest overpredictions and underpredictions addressed in this analysis. While we encourage CMS to explore how outliers affect risk adjustment, addressing these issues will likely require more complex model methods than the approach we evaluated. In addition, more work is needed to understand how the approach presented in this chapter can integrate with other improvements to risk adjustment for MA plans. ■

Endnotes

1 A plan's bid is its estimate of how much it will cost the plan to provide Medicare Part A and Part B services, per enrollee; county benchmarks equal a certain share of the projected average per capita FFS Medicare spending for the county's beneficiaries; a service area is a group of counties for which a plan has agreed to provide services.

2 Much of the bias in the coefficients in the CMS–HCC model is due to Medicare FFS spending at the beneficiary level being skewed such that the distribution has a small share of beneficiaries with very large Medicare spending while most beneficiaries have relatively low levels of spending. The objective of the statistical technique that CMS uses to estimate the model coefficients (weighted least squares) is to find coefficients such that the sum of the squared prediction errors (the difference between a beneficiary's actual costs and predicted costs) is minimized. To achieve this objective, the skewed distribution of beneficiary-level Medicare spending causes the beneficiaries who incur the highest Medicare spending to have a disproportionate effect on the estimated coefficients. The disproportionate effect of these beneficiaries is exacerbated because the weighted least squares estimation method minimizes the squared prediction errors rather than a linear measure of the prediction errors such as absolute value.

3 The method of analysis could be repeated on each of the model segments (six community and institutional segments) independently to maintain revenue neutrality within each model segment. Applying this method to the end-stage renal disease (ESRD) risk model may produce different results if the cost distribution among beneficiaries with ESRD differs from the cost distributions for community and institutional segment populations.

4 The reductions in costs for beneficiaries whose underpredictions exceeded the loss limit was 0.8 × ((actual costs – predicted costs) – loss limit).

5 The increases in costs for beneficiaries whose overpredictions exceeded the gain limit was ((predicted costs – actual costs) – gain limit).

6 Repeating iterations until loss and gain limits changed by less than $1 between iterations is a fairly strict requirement. Original authors of this method determined that only a few iterations were necessary to attain the improvements in model accuracy.

7 The formula for the R^2 is $(1 - \sum(Y_i - \hat{Y}_i)^2 / \sum(Y_i - \bar{Y})^2)$, where Y_i is actual spending for beneficiary i, $\hat{Y}_i$ is predicted spending for beneficiary i, and $\bar{Y}$ is mean spending in our analytic file. The formula for CPM is $(1 - \sum|Y_i - \hat{Y}_i| / \sum|Y_i - \bar{Y}|)$. The only difference between the R^2 and the CPM is that the R^2 has squared terms in the numerator and denominator while the CPM has absolute values in the numerator and denominator.

8 We used half of our analytic sample to estimate the coefficients in both the standard and modified models (the estimation sample), and we used the other half of our analytic sample to obtain other measures of model performance (the evaluation sample). We also used the evaluation sample in an exercise that replicates the method for calculating the R^2 for the standard model and the modified model. The replicated R^2 values round to 0.13 for the standard model and 0.19 for the modified model, the same as the R^2 values from the regressions.

9 Although the modified model produces a better R^2 than the standard CMS–HCC model, we found that full implementation of the reinsurance and repayment policy from McGuire and colleagues would produce an even better R^2 of 0.30.

10 The smaller increase in the CPM relative to the R^2 was expected because the CPM is a linear measure that uses the absolute difference between predicted costs and actual costs, while the R^2 is a quadratic measure that uses the squared differences between predicted costs and actual costs. Because we are addressing the largest differences between predicted costs and actual costs, the effects of reducing overpayments and underpayments will be larger for a measure that includes the squared differences than for a measure that includes the absolute differences.

11 The CMS–HCC model has 83 HCCs while the full DxCG model, the model upon which the CMS–HCC model is based, has 394 HCCs. Research has found that the full DxCG model has an appreciably higher R^2 relative to the CMS–HCC model (Chen et al. 2015). CMS chose to use fewer HCCs in the CMS–HCC model relative to the full DxCG model in response to plan concerns about collecting encounter data for all the HCCs in the full DxCG model (Pope et al. 2004).

References

Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2021. *Report to Congress: Risk adjustment in Medicare Advantage.* Baltimore, MD: CMS.

Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2018. Advance notice of methodological changes for calendar year (CY) 2019 for Medicare Advantage capitation rates, Part C and Part D payment policies and 2019 draft call letter. February 1. https://www.cms.gov/medicare/health-plans/medicareadvtgspecratestats/downloads/advance2018.pdf.

Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2016. Advance notice of methodological changes for calendar year (CY) 2017 for Medicare Advantage capitation rates, Part C and Part D payment policies and 2017 call letter. https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/Advance2017.pdf.

Chen, J., R. P. Ellis, K. H. Toro, et al. 2015. Mispricing in the Medicare Advantage risk adjustment model. *Inquiry* 52.

Frogner, B. K., G. F. Anderson, R. A. Cohen, et al. 2011. Incorporating new research into Medicare risk adjustment. *Medical Care* 49, no. 3 (March): 295–300.

McGuire, T., S. Schillo, and R. C. van Kleef. 2020. Reinsurance, repayments, and risk adjustment in individual health insurance. *American Journal of Health Economics* 6, no. 1: 139–168.

Medicare Payment Advisory Commission. 2020. *Report to the Congress: Medicare and the health care delivery system.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2014. *Report to the Congress: Medicare and the health care delivery system.* Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2012. *Report to the Congress: Medicare and the health care delivery system.* Washington, DC: MedPAC.

Newhouse, J. P., M. Price, J. Huang, et al. 2012. Steps to reduce favorable risk selection in Medicare Advantage largely succeeded, boding well for health insurance exchanges. *Health Affairs* 31, no. 12 (December): 2618–2628.

Pope, G. C., J. Kautter, R. P. Ellis, et al. 2004. Risk adjustment of Medicare capitation payments using the CMS–HCC model. *Health Care Financing Review* 25, no. 4 (Summer): 119–141.