

Fee for Service Adjuster and Payment Recovery for Contract Level Risk Adjustment Data Validation Audits

I. Issue

Medicare Advantage (MA) organizations have asserted that because the Centers for Medicare and Medicaid Services – Hierarchical Condition Category (CMS-HCC) risk adjustment model is calibrated on Fee for Service (FFS) diagnoses that have not been validated by medical record documentation, there is an inaccuracy in payment that should be accounted for in Risk Adjustment Data Validation (RADV) audit recoveries. They have argued that the inaccuracy introduces a systematic bias that results in underpayments which are exacerbated by RADV payment error recoveries. We will refer to this as the “audit miscalibration” assertion.

In 2012, CMS said that it would “apply a Fee-for-Service Adjuster (FFS Adjuster) amount as an offset to the preliminary recovery amount” to be returned under RADV audits. This FFS Adjuster was intended to “account[] for the fact that the documentation standard used in RADV audits to determine a contract’s payment error (medical records) is different from the documentation standard used to develop the Part C risk-adjustment model (FFS claims).” CMS said that “[t]he actual amount of the adjuster will be calculated by CMS based on a RADV-like review of records submitted to support FFS claims data.” That review is now complete.

II. Analytical Approach and Empirical Findings

At a high level, CMS uses the CMS-HCC model to calculate a risk score to represent the relative costliness of each beneficiary as compared to the average beneficiary. The model uses costs and disease diagnoses from FFS claims in an Ordinary Least Squares (OLS) regression framework to associate the diagnoses reported for beneficiaries in one year with FFS costs incurred in the following year. In general, each coefficient estimated by the CMS-HCC model represents the allocation of the FFS costs for an average beneficiary to either a disease or demographic attribute.

As is the case in any OLS model, variance around each parameter estimate determines how precisely the parameter estimate and actual allocations align. Variance can be driven by any number of factors. In this context, the nature of the association between a disease and its cost manifestation across a diverse population of beneficiaries is ostensibly the primary driver. However, other factors drive variation as well. Incorrect information of any type on the claims used in the calibration could influence the variation around the

parameter estimate. The issue of diagnoses unsupported by medical record documentation could be counted in this category.

On average, in the absence of any systematic bias, the parameter estimates and the actual cost allocation are expected to be equal. However, for any given calibration of the CMS-HCC model, because of variance, the parameter estimate of a disease coefficient will almost certainly differ. This difference would be expected to manifest as both greater and lesser than the actual cost allocation. As a consequence, some coefficients would over-estimate the actual cost allocation while others would underestimate it.¹ While the parameter estimates could be described as inaccurate due to variation, the inaccuracy does not necessarily imply a bias.

Medicare uses the CMS-HCC model coefficients to calculate the risk or relative factors. For each beneficiary, the sum of the relative factors is the beneficiary's risk score that is subsequently used to adjust the MA organization's payment for health status. As payment is determined in part by risk score, inaccuracies of the relative factors may or may not affect payment accuracy in the aggregate. Because the impact on any specific relative factor can be positive or negative, its effect does not necessarily create a bias in the risk score or by extension payment in the aggregate. Rather, the impact on an MA organization's payment is dependent on the distribution of its beneficiaries' relative factors and the nature of the inaccuracies.

It is important to clarify that the 'inaccuracy' with respect to the FFS cost comparisons is not isolated to this issue. As just one example, it is well established that the CMS-HCC model under-predicts for high cost enrollees and over-predicts for low cost enrollees.² This is just one among a number of factors that could cause variation in the estimate in comparison to the actual FFS costs. While there is an effect on the variance, there is no bias in the estimate across the program. Accordingly, simply because a mechanism is identified that may lead to payment inaccuracy, it does not, in and of itself, make it inappropriate for Medicare to use risk scores derived in this way as an estimate of a beneficiary's relative costliness due to health status.

As was discussed in the RADV Notice of Final Payment Error Calculation Methodology (February 24, 2012), the first step in evaluating the impact of the audit miscalibration is to do a RADV-like audit on a sample of FFS claims to estimate the prevalence of diagnoses unsupported by medical record documentation. This was done by first mapping every diagnosis on a claim to an HCC. The medical record documentation for each claim

¹ Diagnoses unsupported by medical record documentation in the FFS claims data can result in a coefficient overestimating or underestimating the actual cost allocation of the condition. For example, the inclusion of one such code for a given FFS beneficiary will cause the model to assign some portion of that beneficiary's costs due to other conditions or demographics to the condition in question, but whether this will increase or decrease the coefficient calculated by the model will depend on whether the costs misassigned to the condition for this beneficiary are higher or lower than the costs assigned to this condition for other beneficiaries.

² Report to the Congress: Medicare and the Health Care Delivery System | June 2014, MEDPAC, http://www.medpac.gov/docs/default-source/reports/jun14_ch02.pdf?sfvrsn=0

was than reviewed to confirm there was support for every HCC on the claim. A claim level discrepancy rate was derived for each HCC. The discrepancy rates ranged from 21 to 46 percent.³

In the CMS-HCC regression model, the dependent variable represents the total FFS expenditure for a FFS beneficiary. For each beneficiary, the independent variables include demographic and HCC dummy variables.⁴ For the purposes of this discussion, we will refer to the array of HCC variables associated with a beneficiary's total expenditure as the beneficiary profile. The correctness of the status of an HCC in the profile is determined by the presence of at least one claim that both has a diagnosis that maps to that HCC and is supported by medical record documentation. For this reason, the impacts of the unsupported diagnoses in the claims on the model parameter estimates are driven by the number of HCCs in each beneficiary's profile that do not have at least one claim supported by medical record documentation.

If in a given year, the beneficiary only had one mapping claim for each non-zero HCC in their profile, the probability of it being unsupported would exactly match the claim level HCC discrepancy rate. However, where there are multiple claims with diagnoses mapping to a given HCC, the claim level HCC discrepancy rate must be converted to a beneficiary level discrepancy rate. The beneficiary level discrepancy is determined by adjusting the claim level rate by the number of claims per beneficiary mapping to a given HCC.⁵ The beneficiary level HCC error rates ranged between zero and 15 percent, with the median at 2 percent.

We used the FFS data file built for the 2004-2005 CMS-HCC model calibration as our baseline data. This file includes all the HCCs for a sample of FFS beneficiaries in 2004 (year $t-1$) and those beneficiaries' expenditures in 2005 (year t). We estimated the baseline coefficients by calibrating the CMS-HCC regression model using the baseline data. Using the beneficiary level HCC error rates, we simulated a new FFS data file from the baseline data. HCCs with a status of one simulated to be unsupported by medical record documentation were shifted to zero. Using this new file, we calibrated the CMS-HCC model to estimate HCC coefficients unaffected by FFS claim diagnosis error. We then used the new coefficients to calculate new relative factors from the baseline FFS data set, which, when summed, would result in a normalized risk score that averages to one across all of the beneficiaries in the data set.

In the next step, we calculated risk scores for a sample of MA beneficiaries, both with the baseline coefficients, and also with the new coefficients unaffected by FFS claim diagnosis error. All other things being equal, the difference between the error free and

³ Because these discrepancy rates were derived exclusively from Part B claims, their impact would be expected to be greater than had the rates been derived from both Part A and Part B claims (see technical appendix).

⁴ A dummy variable is valued at one if the status of the dummy is true or zero if the status of the dummy is false.

⁵ A detailed explanation appears in technical appendix.

original risk scores represents the impact of removing FFS error from the calibrating FFS claims.⁶

To measure the effect of the audit miscalibration, we calculated the percent difference between the error free risk scores and the baseline risk scores for every sample MA beneficiary. We then calculated the average difference across all beneficiaries in the sample to get an estimate of the audit miscalibration. A positive audit miscalibration estimate would indicate an underpayment bias, meaning that audit miscalibration results in underpayment to plans. Conversely, a negative value would indicate an overpayment bias, meaning that audit miscalibration results in overpayment to plans.

To account for variance in the beneficiary level HCC discrepancy rates, we simulated the process for fifty iterations to estimate a mean and variance of the audit miscalibration. As shown in Table 1, the simulation estimates the effect of the audit miscalibration to be negative and extremely close to zero.⁷

Table 1. Audit Miscalibration Estimate

Mean (percent difference)	95% Upper Bound	95% Lower Bound
-.08%	-.07%	-.09%

Thus, our study found, within a 95% confidence interval, that audit miscalibration results in a slight (.07% to .09%) overpayment to plans in the aggregate.⁸

⁶ A detailed explanation appears in technical appendix.

⁷ A more detailed explanation of the simulation can be found in the technical appendix.

⁸ Prior to the 2012 announcement where CMS said it would apply a FFS adjuster, the agency investigated the industry assertion that FFS diagnoses errors on claims would reduce risk scores for a given model calibration. Using claim level discrepancy rates from a RADV-like audit, CMS measured the nominal reduction in risk scores from turning off HCCs in the risk score calculation at the claim level discrepancy rate. The estimates of the reductions ranged between 4.8% and 8.1%. Because this analysis used claim level rather than the beneficiary level discrepancy rates, which, as explained above and in the technical appendix, are not the appropriate rates, it likely greatly overstated the effects of the diagnoses errors. Additionally, the earlier analysis did not analyze the effect of FFS diagnosis error on the CMS-HCC model. In contrast to the earlier analysis, the analysis explained in this study and the technical appendix re-calibrates the CMS-HCC model on data corrected with beneficiary level discrepancy rates, calculates new normalized relative factors, calculates new error free risk scores, and compares the new scores to error affected risk scores, comprehensively measuring the effect of FFS diagnosis error on the CMS-HCC model.

III. Conceptual Issues with the FFS Adjuster

We have empirically demonstrated that while there is significant diagnosis error in FFS claims, the impact is less than one percent on average and in the favor of the plans. While a particular HCC's relative factor may have inaccuracy attached to it, the fact that the relative factors are summed across each enrollee's HCCs and then across a plan's enrollment, leads the inaccuracies to mitigate each other due to offsetting effects.⁹ Additionally, because the CMS-HCC model regresses FFS costs on a beneficiary's HCC profile at the beneficiary rather than the claim level, the impact of diagnosis error on the claims is reduced exponentially by the number of claims that map to a given HCC.

Payment standards require that submitted diagnoses must be supported by medical record documentation. RADV audits are used to recover payments based on diagnoses that are not supported by medical record documentation. If a payment has been made to an MA organization based on a diagnosis code that is not supported by medical record documentation that entire payment is in error and should be recovered in full. RADV audits do not address issues with the accuracy of payments based on diagnosis codes that are supported by medical record documentation.

Consequently, an adjustment to RADV recoveries to remedy payment accuracy concerns is inappropriate and problematic. As discussed, the impact of any audit miscalibration is tied to the characteristics of a plan's enrollment. Due to those characteristics, some plans may be overpaid while others are underpaid. This would make the adjustment arbitrary and would introduce inequities between audited and unaudited plans by only correcting the payments made to audited plans. Additionally, if differences in the disease profiles of the MA population relative to the FFS population put a higher proportion of beneficiaries on overpaid HCCs, an adjustment could exacerbate improper payment. Accordingly, we conclude that it is inappropriate to apply the FFSA to RADV recoveries to mitigate any payment inaccuracy. If a payment inaccuracy due to audit miscalibration bias exists, which we believe unlikely for the reasons already discussed, an appropriate remedy would be an adjustment to original payments at the MA organization level.

IV. Conclusion

We have now completed the "RADV-like review of records submitted to support FFS claims data" that was announced in February 2012 and have found that errors in FFS claims data do not have any systematic effect on the risk scores calculated by the CMS-HCC risk adjustment model, and therefore do not have any systematic effect on the payments made to MA organizations. While our empirical findings show significant diagnosis error in FFS claims, we were unable to find a material impact to the model calibration indicating a payment bias. Further, the plan level impact of the audit

⁹ As a statistical phenomenon, certain individual HCCs with measurement error may be subject to downward biases. However, this will result in upward biases to other HCCs and demographic factors. Across HCCs, these biases are likely to offset. The degree of offset is an empirical question.

miscalibration is a function of each plan's enrollee HCC distribution. The net impact on plan payment is determined by the share of enrollees with HCC coefficients that are lower or higher due to audit miscalibration. This is consistent with the effects on payment of normal variation in any of the population characteristics included in the CMS-HCC model. As a consequence, an adjustment to payment would arbitrarily reduce improper payment for some plans while increasing it for others.

Because our study suggests that diagnosis error in FFS claims data does not lead to systematic payment error in the MA program and because we believe it would be inequitable to correct any systematic errors in the payments made to audited plans only, we no longer believe it is appropriate to include an FFS Adjuster in any RADV extrapolated audit methodology.