Medicare Advantage RADV FFS adjuster: White paper

Commissioned by America's Health Insurance Plans (AHIP)
August 23, 2019

Rob Pipich, FSA, MAAA

Executive Summary

The Centers for Medicare and Medicaid Services (CMS) issued a proposed rule[1] on November 1, 2018, which contained provisions regarding risk adjustment data validation (RADV) audits. In particular, this proposed rule removed what is known as the fee-for-service (FFS) adjuster, which is a mechanism for adjusting RADV audit recoveries to ensure actuarial equivalence between FFS and MA payments. Actuarial equivalence is required by law.[2] Based on the analysis described in this white paper, we determined:

- A FFS adjuster, or other similar adjustment, is necessary to ensure actuarial equivalence between payments to Medicare Advantage Organizations (MAOs) and payments under Medicare FFS.
- CMS analyzed the difference between two calibrations of the CMS Hierarchical Condition Category (HCC) model to investigate what it referred to as "audit miscalibration."[3] CMS normalized the revised model inconsistently within the context of a FFS adjuster or a RADV audit; therefore, its technical analysis cannot appropriately be used to conclude a FFS adjuster is not required.
- CMS underestimates the level of diagnosis coding errors present in FFS claims data. Notably:
  - CMS assumes diagnosis coding errors are independent from each other, which materially understates HCC error rates in FFS.
  - CMS uses an average number of claims per HCC in its estimation of error rates rather than a distribution of the number of claims, which materially understates HCC error rates in FFS.
  - CMS excludes claims that do not have medical records or necessary documentation available, which also understates the HCC error rates in FFS relative to RADV audit procedures.

This white paper discusses and supports our findings that a FFS adjuster is required in RADV audits. The CMS technical analysis excluded simulated unsupported diagnoses in the calibration of the CMS-HCC model, but included them in the normalization of the model. CMS should have excluded unsupported FFS diagnoses in all steps of creating the CMS HCC model to properly address the question of whether a FFS adjuster is required in RADV audits. This paper shows, had CMS excluded unsupported diagnoses from all steps, their analysis would have confirmed a FFS adjuster is required.

[1] Medicare and Medicaid Programs; Policy and Technical Changes to the Medicare Advantage, Medicare Prescription Drug Benefit, Program of All-inclusive Care for the Elderly (PACE), Medicaid Fee-For-Service, and Medicaid Managed Care Programs for Years 2020 and 2021, 83 Fed. Reg. 54982 (2018). Retrieved December 20, 2018, from https://www.gpo.gov/fdsys/pkg/FR-2018-11-01/pdf/2018-23599.pdf.

[2] Title 42 U.S. Code § 1395w–23(a)(1)(C)(i).

[3] CMS coins the term "audit miscalibration" in its FFS adjuster executive summary. Retrieved December 20, 2018, from https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-Risk-Adjustment-Data-Validation-Program/Other-Content-Types/RADV-Docs/FFS-Adjuster-Excecutive-Summary.pdf. The proposed rule describes a similar concept. 83 Fed. Reg. 55041 (2018).

The key items presented in this white paper include:

- An explanation for why a FFS adjuster is required in a RADV audit to maintain actuarial equivalence, as required by statute and confirmed in *UnitedHealthcare Ins. Co. v. Azar*[4].
- A simplified numeric example demonstrating the argument described in the prior bullet. This is an example expanded upon from an example created by CMS.
- A summarized description of CMS's detailed technical analysis and an explanation of why we believe the methodology does not support the removal of a FFS adjuster.
- An adjusted version of the CMS analysis using a consistent set of diagnoses throughout the entire analysis showing why we believe a FFS adjuster or similar adjustment mechanism is necessary.
- A discussion of CMS's development of the Medicare FFS HCC error rates, which we conclude results in a significant understatement of the HCC error rates and therefore should not be used in assessing the magnitude of the FFS adjuster.

PURPOSE OF THIS STUDY

The purpose of this study is to evaluate the CMS conclusion that a FFS adjuster is not appropriate; it is not to determine the appropriate amount of a FFS adjuster. The study shows that using CMS' methodology and data but adjusting for certain issues with that methodology, as described in this paper, leads to a conclusion that a FFS adjuster is required and is significantly greater than zero. As described in various sections of this paper, including those titled (a) 'CMS underestimated error rates for HCCs – Overview', (b) 'CMS underestimated error rates for HCCs – Is the sample size sufficient?', (c) 'Technical analysis - Model and data selection', and (d) 'Conclusion', further study of error rates is necessary to determine the true magnitude of a FFS adjuster.

This study uses CMS published assumptions, methodology, and data, and identifies multiple significant issues in CMS assumptions and methodologies. We did not attempt to identify all potential issues. We make no judgment about the appropriateness of other methodologies that could be used to determine an appropriate FFS adjuster. Depending on other potential issues and alternative assumptions and methodologies used, other valid analyses may lead to reasonable FFS adjusters that are outside the ranges considered in this paper. However, we have not been able to conceive of a reasonable methodology that would lead to the conclusion a FFS adjuster is unnecessary.

BACKGROUND AND DEFINITIONS

MAOs are paid, in large part and with certain adjustments, based upon the expected cost of the individual beneficiaries who enroll in the MAO's plans had those beneficiaries received benefits through the Medicare FFS program. Generally, CMS uses a risk adjustment system to multiply a fixed monthly capitation payment times a beneficiary-specific risk score to adjust payments to MAOs based on health status. That approach to determining the capitation payment results in higher payments to MAOs for less healthy beneficiaries and lower payments for healthier beneficiaries.

Title 42 of the United States Code § 1395w–23(a)(1)(C)(i)[5] states that the risk adjustment mechanism used by CMS should be implemented in a manner that achieves actuarial equivalence between Medicare FFS and Medicare Advantage (MA). CMS recognized this requirement in its February 24, 2012, notice[6], which set forth the methodology for RADV audit recovery calculations. The notice acknowledged that the CMS HCC risk score model is developed based upon diagnoses from FFS claims, including those not supported by medical records. Therefore, if a RADV audit removes unsupported diagnoses from an MAO's risk score calculation, the MAO must be allowed the same level of unsupported diagnoses as FFS

4 330 F.Supp.3d 173 (D.D.C. 2018) (Collyer, J.), *appeal docketed*, No. 18-5326 (D.C. Cir. Nov. 14, 2018).

5 Available at https://www.law.cornell.edu/uscode/text/42/1395w-23.

6 CMS (February 24, 2012). Notice of Final Payment Error Calculation Methodology for Part C Medicare Advantage Risk Adjustment Data Validation Contract-Level Audits. Retrieved December 20, 2018, from https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/recovery-audit-program-parts-c-and-d/Other-Content-Types/RADV-Docs/RADV-Methodology.pdf.

in order to maintain actuarial equivalence. Failing to do so would result in CMS paying less, on average, for an identical beneficiary under the MA program than under the FFS program, violating the principle of actuarial equivalence.

To avoid confusion throughout this paper, we define a few terms. The term "calibrate," as it applies to the HCC model, is often loosely used to refer to both the process where CMS *calibrates* the HCC model and then *normalizes* the model. In this white paper, we use the term *calibrate* to refer to the application of a least squares regression to calculate the relative cost of medical conditions and demographic indicators included in the HCC model. We use the term *normalization* to refer to the process by which CMS ensures that the HCC model, when applied to the FFS population, results in a 1.0 average risk score.

ANALYSIS AND RESULTS

In the February 24, 2012 notice, CMS acknowledged the need for a FFS adjuster and included it in the RADV audit procedures. CMS is now proposing to remove the FFS adjuster. The CMS proposal to remove the FFS adjuster is primarily supported by a technical analysis[7] showing that calibrating the HCC model using a data set containing all diagnoses versus only supported diagnoses (i.e., diagnoses supported by medical records) does not materially impact overall MAO payment levels. CMS argues this occurs because of the normalization process CMS uses to ensure that the average risk score for the FFS population is 1.0. However, it appears that CMS performed the normalization process by including unsupported diagnoses that should have been excluded. The result is that the portion of the CMS analysis intended to represent a scenario without unsupported diagnoses does not, in fact, remove the unsupported diagnoses.

The CMS Technical Appendix[8] did not provide all the details surrounding the technical calculations in the CMS analysis, and so to initially confirm our understanding of what CMS did, we successfully reproduced the CMS technical analysis described above. We reproduced the CMS analysis using both the data CMS released in March 2019 to support its technical analysis and using the 2014/2015 Limited Data Set 5% Samples (5% Samples).[9] CMS subsequently (June 2019) published an 'Addendum to the Fee-For-Service Study' (Addendum)[10], which included many of the previously missing technical calculation details and certain CMS SAS code; we verified that the CMS implementation of the process described in the technical appendix was not materially different from our reproduction of the CMS analysis.

The CMS analysis includes certain simplifying assumptions that result in materially understated FFS HCC error rates. The CMS simulations used those understated FFS HCC error rates. We analyzed several variations of the CMS technical analysis: (a) excluding the unsupported diagnoses (from not only the HCC model calibration process, but also from the normalization process), (b) calculating HCC level error rates based on the CMS claim level error rates and the actual distribution of claims per beneficiary (as opposed to the average across all beneficiaries), and (c) testing several levels of HCC error rates. We used the CMS error rates and methodology with an adjustment for the normalization process and the actual number of diagnoses per beneficiary (rather than the average). Under this approach, we calculated a FFS adjuster using claim level error rates, actual distributions of the number of diagnoses (assuming full *independence*[11]), and an HCC error rate of 33% (assuming full *dependence*[12]), in addition to several

---

[7] CMS (October 26, 2018). RADV Resources. Retrieved December 20, 2018, from https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-Risk-Adjustment-Data-Validation-Program/Resources.html.

[8] Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-Risk-Adjustment-Data-Validation-Program/Resources.html.

[9] The 5% Samples are Limited Data Sets made available by CMS and we utilized the particular files that contain approximately 5% of Medicare member's FFS claims. Additional information is available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/StandardAnalyticalFiles.html

[10] Retrieved June 26, 2019 from https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-Risk-Adjustment-Data-Validation-Program/Other-Content-Types/RADV-Docs/RADV-Provision-CMS-4185-N4-Data-Release-June-2019.zip.

[11] Independence, in this context, means coding errors on individual claims are not related to coding errors on other claims.

[12] Dependence, in this context, means coding errors on claims are made in the same way for all claims for a particular HCC for each beneficiary.

scenarios in between. This approach resulted in estimated values of a FFS adjuster[13] between 8% and 21%. For perspective, 8% of federal payments to MAOs exceeds $16 billion and 21% exceeds $42 billion per year,[14] the majority of which are risk-adjusted.

A FFS adjuster, based on CMS's data modified to reflect reasonable error rates using an adjusted methodology (e.g., adjusts for the normalization process, the distribution of claims, and claim independence) likely lies somewhere between the two endpoints, 8% and 21%. We also note that CMS clarified in the June 2019 Addendum that they "…excluded claims where providers refused to submit medical records, or did not provide sufficient documentation." Although we do not have the information to evaluate the impact of these exclusions on the error rates, this exclusion is inconsistent with the RADV audit process. Properly including these unsupported diagnoses in the calculation of error rates would increase the magnitude of a FFS adjuster from the figures described in this paper.

As noted above, we make no judgment about the appropriateness of other methodologies that could be used to determine an appropriate FFS adjuster. Depending on other potential issues and alternative assumptions and methodologies used, other valid analyses may lead to reasonable FFS adjusters that are outside the range considered in this paper.

The magnitude of a FFS adjuster is highly sensitive to the specific HCC error rates used in the analysis, and the HCC error rates in the CMS analysis are highly sensitive to both the use of an average number of claims (versus a distribution of the number of claims) within an HCC and how independent the coding of one claim is to the next.

Further analysis must be completed to calculate an accurate FFS adjuster. In any case, the range is wide and even the bottom end is material and significant.

We conclude that not applying a FFS adjuster in a RADV audit, as proposed by CMS, would violate actuarial equivalence. Additionally, applying a FFS adjuster based on the HCC error rates in the CMS Technical Appendix would also violate actuarial equivalence because the HCC error rates CMS uses are biased. A FFS adjuster must be developed consistent with the intended application to ensure actuarial equivalence.

I, Rob Pipich, am a Member of the American Academy of Actuaries and I meet the Qualification Standards of the American Academy of Actuaries to render the actuarial opinions expressed herein.

Introduction

The issues involved in Medicare risk scores, RADV audits, and actuarial equivalence are complex. We organize this white paper to facilitate a simpler way to understand the issues. The executive summary above provides an overview of our analysis and findings. The remaining sections describe our analysis in more detail and provide support for our findings. The following is a list of the topics in the order we address them:

- Background
- Actuarial equivalence requires a FFS adjuster in RADV
- CMS technical analysis should not include unsupported FFS diagnoses
- CMS underestimated error rates for HCCs
- A CMS example demonstrating the need for a FFS adjuster

[13] We define the FFS adjuster as the percentage reduction to a risk score based upon claim diagnoses to move to a medical record diagnosis basis for a FFS population. We calculated this percentage including beneficiaries with no HCCs and beneficiaries with one or more HCCs. When applying a FFS adjuster, care must be taken to apply it to the correct population, as the difference between the two definitions is significant. If this adjuster is applied to only beneficiaries who are RADV-eligible under the current CMS rules, the adjuster would need to be grossed up to apply only to that population.

[14] Based on $204.7 billion in 2017 Part C federal spending. See HHS FY 2017 Budget in Brief - CMS – Medicare, available at https://www.hhs.gov/about/budget/fy2017/budget-in-brief/cms/medicare/index.html.

- An expanded example incorporating normalization and RADV audits
- Discussion of our technical analysis, which mirrors the CMS analysis
- Additional context and considerations surrounding the HCC risk model and a FFS adjuster
- Conclusion
- Appendices of additional charts and examples

Background

MAOs are paid fixed per beneficiary amounts to deliver care to Medicare beneficiaries. These fixed amounts are calculated based upon a combination of amounts MAOs submit to CMS in the annual bid process and the projected health status of each beneficiary as determined from their actual diagnoses and demographic information. While the complexities of the bid process are outside the scope of this paper, the majority of funding from CMS to MAOs is calculated by multiplying the plan bid amount at a 1.0 risk score times the actual risk score of the beneficiary. As a result, the actual beneficiary risk scores are a key determinant of total revenue for MAOs.

Risk scores are calculated based upon diagnosis information from claims data using the CMS HCC model. Generally, more diagnoses result in higher payments by triggering more HCCs. It is worthwhile to note that not all diagnoses map to an HCC and coding the same HCC more than once for an individual does not impact the risk scores.

CMS calculates the dollar amount each HCC is worth in the CMS HCC model utilizing a weighted least squares regression, with certain constraints,[15] based on one year of FFS *diagnosis* data from claims and the following year's FFS *claims cost* data. In essence, the dollar amount each HCC is worth, divided by the overall average claims cost for a FFS beneficiary, is referred to as the coefficient for each HCC. The steps thus far are typically referred to as "c*alibration*." To *normalize* the model to a 1.0 risk score for the FFS population, CMS calculates an average risk score for the FFS population and then divides all model coefficients by that average FFS population risk score. Additional details regarding the creation of the CMS HCC model can be found in "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model," published in *Health Care Financing Review*.[16]

The term "calibrate," as it applies to the HCC model, is widely used to refer to both the process where CMS *calibrates* the HCC model and then *normalizes* the model. In this white paper we clarify and distinguish the terms and use *calibrate* to refer to the application of a least squares regression to calculate the relative cost of medical conditions included in the HCC model. We use the term "*normalization"* to refer to the process by which CMS ensures that the HCC model, when applied to the FFS population, results in a 1.0 risk score on average.

After diagnoses have been reported and CMS issues final payments to MAOs based upon the final diagnoses, CMS then performs RADV audits on a selected set of MAOs. CMS's stated intent for RADV audits is to validate the accuracy of risk-based payments by validating the diagnoses, through medical records, submitted by MAOs that map to an HCC for payment. Conceptually, through these RADV audits, CMS intends to recover overpayments made to MAOs.

---

[15] The constraints are technical in nature, such as disallowing negative coefficients.

[16] Pope, G.C., Kautter, J., Ellis, R.P. et al. (2004). Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financing Review*, Summer 2004, Vol. 25 No. 4. Retrieved December 20, 2018, from https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/HealthCareFinancingReview/downloads/04summerpg119.pdf.

As described in the notice dated February 24, 2012[17], RADV audits, in general simplified terms, involve:

1. Excluding end-stage renal disease (ESRD) and hospice beneficiaries as well as any beneficiary not continuously enrolled from January of the diagnosis year to January of the payment year and who does not have an HCC.
2. Ranking beneficiaries in each MA contract by risk score and dividing them into three equal groups.
3. Sampling 67 beneficiaries from each group.
4. Requesting and auditing medical records from the MAO for each HCC recorded among the sampled beneficiaries.
5. Calculating a "payment error" based on the difference in the original payment and the RADV-audit-adjusted payment.
6. Calculating a 99% confidence interval (CI) for the annual payment error per MA contract.
7. Selecting the lower bound of the CI and, if it is above zero, reducing it by the FFS adjuster.
8. Extrapolating (for recovery) the result of step 7 to every RADV eligible beneficiary in the contract if the result of step 7 is a positive value.

CMS also stated the following in the February 24, 2012 notice for the rationale for a FFS adjuster:

> "The FFS adjuster accounts 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)."

On November 1, 2018, CMS published the proposed rule proposing to eliminate the FFS adjuster, with a comment deadline of December 31, 2018. Subsequently, CMS extended the comment deadline for stakeholders to April 30 after announcing it would publish additional data. CMS again extended the deadline to August 28, 2019, publishing the programming code and additional data from 50 new simulations that CMS ran. Milliman obtained and evaluated the additional data. The proposed rule's provisions on RADV is the subject of this white paper.

Actuarial equivalence requires a FFS adjuster in RADV

If CMS applies different standards for determinations of diagnoses under Medicare FFS and MA, the required actuarial equivalence is not achieved. In its proposed rule, CMS proposes to apply the claims diagnoses to Medicare FFS and the medical record diagnoses to MAOs under RADV audit. This approach will not generate actuarially equivalent results without an adjustment to account for the difference between the claims diagnoses and the medical record diagnoses present in the FFS data (e.g., a FFS adjuster).

The Secretary of the U.S. Department of Health and Human Services (HHS) implemented the CMS HCC risk score model under authority granted by Title 42 U.S. Code § 1395w–23(a)(1)(C)(i), underline and bold added for emphasis:

> "(C) Demographic adjustment, including adjustment for health status
>
> (i) In general The Secretary shall adjust the payment amount…for such risk factors as age, disability status, gender, institutional status, and such other factors as the Secretary determines to be appropriate, <u>including adjustment for health status under paragraph (3), so as to **ensure actuarial equivalence**</u>. The Secretary may add to, modify, or substitute for such adjustment factors if such changes will improve the determination of actuarial equivalence."

[17] During the comment period for the proposed rule, CMS released a notice revising the RADV audit procedures for 2014. Since these procedures are not finalized, will be subject to the final rule, and are not included in the CMS analysis accompanying the proposed rule, we do not comment on the 2019 notice in this paper.

As stated by CMS in the February 24, 2012 notice, the documentation standard used to determine payment errors under a RADV audit of an MAO is medical records, but the documentation standard used to develop the HCC model is FFS claims data. The introduction of this different documentation standard violates actuarial equivalence unless a FFS adjuster is included.

In *UnitedHealthcare Ins. Co. v. Azar*,[18] the court ruled that both (a) Title 42 U.S. Code § 1395w–23(a)(1)(C)(i) requires the Secretary to implement a risk adjustment program that effectuates *actuarially equivalent* risk adjustment of payments between the FFS and MA programs, and (b) the varying documentation standards violate actuarial equivalence.

Further, CMS itself, in internal documents released in response to a Freedom of Information Act (FOIA) request, agrees and states: “We think this approach makes sense and from a technical point of view is the right thing to do,”[19] in reference to including a FFS adjuster to address the issue of differing coding standards.

CMS technical analysis should not include unsupported FFS diagnoses

CMS included both supported and unsupported diagnoses in the technical analysis it described as simulating HCC model creation with only supported diagnoses. In effect, the CMS technical analysis compared a model created with all diagnoses to another model created with all diagnoses, effectively making the analysis irrelevant to the discussion of whether or not a model calibrated and normalized using only supported diagnoses would produce different payments to MAOs. Stated differently, the CMS analysis did not serve to address the question of whether or not a FFS adjuster is necessary. We discuss specific assertions and explanations put forth by CMS in the ‘Technical Analysis’ section, below. The remainder of this section focusses on a conceptual discussion, followed by examples, both of which clearly demonstrate the need for a FFS adjuster and that the CMS technical analysis should not include unsupported FFS diagnoses.

The CMS technical analysis was put forth to demonstrate that MAO payments do not materially change based upon calibrating the CMS HCC model, including or excluding unsupported HCCs. However, the calibration of the model is only a portion of the issue, and for the other portion of the issue, which is normalization, CMS did not exclude the simulated unsupported HCCs.

We summarize the CMS description of its process as:

1. Calibrate the HCC model utilizing the original uncorrected data set (where the uncorrected data set includes unsupported diagnoses).
2. Normalize this HCC model using the original uncorrected data set to achieve a 1.0 risk score in total.
3. Calculate claims-level error rates using the FFS Comprehensive Error Rate Testing (CERT) data.
4. Convert the claims-level error rates into HCC-level error rates.
5. Utilize the HCC error rates to simulate the removal of unsupported diagnoses from the original uncorrected data set to produce a simulated corrected data set.
6. Calibrate the HCC model utilizing the simulated corrected data set.

[18] 330 F.Supp.3d 173 (D.D.C. 2018) (Collyer, J.), *appeal docketed*, No. 18-5326 (D.C. Cir. Nov. 14, 2018). Retrieved December 1, 2018, from https://dlbjbjzgnk95t.cloudfront.net/1046000/1046226/https-ecf-dcd-uscourts-gov-doc1-04516758487.pdf.
[19] See Appendix E below. Acquired from: DOCKET 44. UNITED V PRICE NO. 1:16-CV-00157-RMC.

7. Normalize this HCC model using the original uncorrected data set to achieve a 1.0 risk score in total.[20]
8. Apply both models to a sample MAO data set and compare the resultant risk scores.

The calibration of the HCC model utilizes a weighted least squares regression (see the Statistical Background section below for more details) to determine how the risk score coefficient for each HCC relates to the coefficients for other HCCs. For example, calibration might determine that the coefficient for HCC 1 is 10% higher than HCC 2, 25% higher than HCC 3, etc. The calibration step does not determine the final level of the coefficients.

It is the normalization step that determines the final level of the coefficients for each HCC. Specifically, CMS applies the calibrated risk score model to the uncorrected FFS data set, and divides all the coefficients by the resulting risk score to ensure that the final normalized model produces an average risk score of 1.0 for the FFS population. CMS used the uncorrected data set to normalize both the HCC model that was calibrated with the uncorrected data set and the HCC model that was calibrated with the simulated corrected data set.

The calibration step is not significant in the context of determining the overall risk score. It simply adjusts the relative value of each HCC. The normalization step is critical because it scales how much each HCC counts in determining an overall risk score. As mentioned, CMS utilized the uncorrected data set to normalize both HCC models, so neither model reflects the removal of the simulated unsupported diagnoses.

Stated differently, CMS removed the simulated unsupported diagnoses for the calibration step and then immediately put them back into the analysis for the normalization step. When CMS compared the MA risk scores produced by the two different models, it was really calculating the effect of MA having a slightly higher incidence of certain HCCs than FFS and a slightly lower incidence of others (which is the small difference it identified). In CMS's technical analysis comment, CMS references a potential difference of this sort and discards it as possible but immaterial. The CMS analysis does not compare the effect of calibrating and normalizing the model with and without unsupported HCCs, which is critical for calculating a FFS adjuster.

An example, developed by CMS, illustrating this concept is included in Appendix A. We expanded the example explicitly to include the impact of model normalization in the section entitled "Example demonstrating actuarial equivalence is violated."

There also appears to be an inconsistency in the CMS Technical Appendix. Specifically, CMS describes a proper procedure but then uses a different procedure in practice. On page 12 of the appendix, CMS states (emphasis added):

> "Although fundamentally based on expenditures, the regression is adjusted such that the HCC and demographic factors will provide an average risk score of one <u>on the calibrating FFS dataset</u>."

As described on the next page of CMS's Technical Appendix (emphasis added):

> "We then estimate the CMS-HCC model on the <u>simulated corrected data</u>. In the next step, we take the new coefficients and apply them on the <u>original FFS data set</u>, normalizing a new set of relative factors to one."

[20] In documents such as rate announcements and proposed rules on risk scores, CMS describes a process of creating the CMS HCC risk model as including a step to divide dollar-based HCC coefficients by a total denominator year predicted cost. The Technical Appendix does not describe this step, but does describe normalization. As we interpret the CMS Technical Appendix, the normalization step is comparable to the denominator year adjustment. This understanding is supported by additional details provided by CMS in the June Addendum.

Because CMS has normalized back to the “original FFS data set” and not the “simulated corrected data,” which was the “calibrating FFS dataset,” CMS effectively added the simulated unsupported diagnoses back into the data set, which sets the documentation standard back to a claims diagnosis basis. Thus, the CMS analysis measured a model calibration difference rather than addressing the question of whether a FFS adjuster is required in RADV audits.

CMS underestimated error rates for HCCs

OVERVIEW

CMS established HCC error rates for the purpose of evaluating a FFS adjuster utilizing data and methodologies that led to underestimation of the HCC error rates. In the Technical Appendix, CMS recognizes certain shortcomings in the calculation of error rates and the data used to calculate the error rates.

Utilizing the potential range of HCC error rates from the CMS analysis that would result from alternative assumptions regarding the degree of independence of claims-level error rates, we estimate that CMS significantly understated the HCC error rates. Specifically, CMS utilized an aggregate HCC error rate of 2% when the true error rate, based on CMS data and varying the degree of dependence, is likely to be between 12% and 33%.

Appropriate testing of FFS data to support the calculation of an HCC error rate must be performed to properly calculate the magnitude of a FFS adjuster. In particular, all claims for a sample of beneficiaries must be used rather than a sample of claims from a wide array of beneficiaries that are converted to a beneficiary basis. Claims must not be excluded simply because the provider did not provide sufficient medical records or documentation, because a RADV audit would include such claims and count them as unsupported diagnoses (i.e. errors). Further, claims from all settings of care should be used with an appropriate sample size. Stratified sampling by HCC combined with oversampling for low frequency HCCs may be an appropriate method to reduce the required sample size.

CLAIMS CODING ERROR INDEPENDENCE

The CMS Center for Program Integrity (CPI) “performed a RADV-like review on the CERT data,” which included 2008 outpatient FFS diagnosis data. Claims were only included if they had diagnoses that mapped to HCCs; 8,630 unique claims were included, which is a relatively small total sample size given the large number of diagnosis codes and HCCs.

While CMS stated that it used “RADV-like review” procedures, CMS deviated from RADV procedures in several important ways. CMS did not include claims for which providers did not provide sufficient medical record support. Further CMS did not review all claims for individual beneficiaries; rather, CMS reviewed and calculated error rates on individual outpatient claims. An audit using all claims mapped to an HCC for a representative sample of individual beneficiaries is necessary to properly estimate the HCC error rate for the Medicare FFS program.

Diagnoses can be coded by different providers in different settings. Coding of a single supported diagnosis that maps to a particular HCC is sufficient to include that HCC for a beneficiary. As such, accurate estimation of HCC error rates must be completed by reviewing all the claims with diagnoses that trigger an HCC for an individual beneficiary and determining whether or not at least one diagnosis is supported by the medical record.

Many coding errors are not independent from one claim to the next. CMS’s approach ignores any correlation between coding errors, effectively assuming that providers randomly make coding errors without regard to errors they have made in the past. We believe it is more likely that a provider or medical coder would tend to make similar errors from one claim to the next based upon their work habits, training, office practices, and by looking at their own prior diagnosis coding when coding a subsequent claim; thus errors would be correlated, at least to some degree. The assumption that providers code randomly must hold to assume independence.

For example, for a beneficiary with Major Depressive, Bipolar, and Paranoid Disorders (HCC 55),[21] CMS calculated a claims-level coding error rate of about 50%, the same probability of flipping heads in a coin toss. CMS further calculated a beneficiary with HCC 55 is likely to have about six claims per year with diagnoses mapping to HCC 55. CMS then assumed each claim is independent, as flips of a coin are independent. Under this assumption of independence, we would statistically expect three codes for an average beneficiary to be supported and three codes to be unsupported. Under this scenario where providers behave randomly (like a coin flip), it would be extremely unlikely to have six coding errors on six visits (like flipping heads six times in a row). This independence assumption can be expected to result in HCC-level error rates that are significantly lower than if providers or medical coders make errors that are related to each other, perhaps from copying diagnoses from a prior visit or from particular personnel repeatedly making the same type of error.

Nevertheless, in calculating HCC error rates CMS has assumed independence of errors among claims. CMS assumes that each claim is equally and independently likely to have an unsupported diagnosis coded. As such, CMS raises the probability of an error on a single claim to the power of the average number of claims. In our example with HCC 55, CMS assumes the probability of that error occurring six times for the same beneficiary is 0.5 ^ 6 = 1.6%. Another scenario where the claims-level error rate is 50% for beneficiaries with HCC 55 can be illustrated simply by considering two beneficiaries. Assume both beneficiaries have HCC 55, visited a provider six times, and have had HCC 55 for several years. Beneficiary A's provider reviewed the patient history and copied support for HCC 55 in the electronic medical record from prior visits and pasted it in the medical records again for the current year, but Beneficiary B's provider continued treating the patient without rerecording the support in the medical record. In this example, Beneficiary A has six supported diagnoses and Beneficiary B has zero, resulting in a claims-level error rate of 6 / 12 = 50%. The HCC error rate is also 50% (1 / 2 = 50%). The assumption of independence significantly reduces the HCC error rate. In the example illustrated here and looking solely at the issue of independence, the true HCC error rate can be expected to be between 1.6% and 50%, depending upon the specific coding patterns of the providers and medical coders involved.

A provider's work habits, job training, and office operating procedures all lead to an increase in the degree of dependence in coding errors. For example, if a particular provider's office has a gap in its training of medical coders around coding diagnoses that map to HCC 55, those coders are likely to repeatedly make the same mistake. This could lead to every beneficiary who is treated by the office having the same coding error for every claim. This leads to the same result described in the previous paragraph's example. Under this assumption, each beneficiary has a 50% chance of an HCC error being recorded. Again, the true HCC error rate can be expected to be between this scenario and the full independence error rate, that is, between 1.6% and 50% for HCC 55.

THE AVERAGE NUMBER OF CLAIMS PER BENEFICIARY CANNOT BE USED TO TRANSLATE TO A BENEFICIARY- LEVEL ERROR RATE

Using the average number of claims per beneficiary materially understates error rates when translating claims-level error rates to beneficiary-level error rates. We describe our approach for adjusting for this issue in the 'Technical analysis: Model and data selection' section of this paper, below.

The CMS technical analysis uses the average number of claims per beneficiary to convert a claims-level error rate to a beneficiary-level HCC error rate. Ignoring the issue with independence, as discussed above, failing to account for the distribution of the number of claims per beneficiary within an HCC will bias the error rate downward from the true value.

[21] HCC 55 included 448 claims in the 2008 FFS CERT sample, a number likely to be credible to calculate the error rate on claims. The exact error rate CMS calculated was 51.80%. Further, CMS estimated 6.1 visits per year per beneficiary with a diagnosis code mapping to HCC 55.

Some beneficiaries will have more claims than the average and some will have fewer. The approach CMS uses applies an exponent, which represents the average number of claims per HCC, to claims-level error rates that are below 1.0. As the number of claims increases, adding an additional claim does not materially change the assumed HCC-level error rate. However, at a lower number of claims per HCC, each additional claim does make a material difference.

Consider a continuation of the HCC 55 example from above. CMS assumes an average claims error rate of 50% with an average of six claims per beneficiary. If there are two beneficiaries with HCC 55 and one has two claims while the other has 10, the HCC error rate (assuming independence for simplicity only) is 0.5 ^ 2 = 0.25 for the first beneficiary and 0.5 ^ 10 = 0.0001 for the second beneficiary. Averaging these error rates yields an average HCC error rate of 0.125. A similar calculation utilizing an average number of claims for all beneficiaries yields an average error rate of 0.5 ^ 6 = 0.016 for each beneficiary. The true average error rate in this example is nearly eight times higher than the error rate calculated using an average number of claims per beneficiary.

SENSITIVITY OF A FFS ADJUSTER TO ERROR RATES

The results of the CMS study are very sensitive to the specific error rates used in the analysis. The error rates are highly sensitive to how independent the coding of one claim is to the next as well as to the distribution of the number of claims with a diagnosis mapping to a particular HCC. We performed sensitivity analyses and present the FFS adjusters we calculated when assuming (a) full independence with an average number of diagnoses per beneficiary, (b) full independence with a distribution of the number of diagnoses per beneficiary in the 2014 5% Sample, (c) complete dependence, and (d) 25%, 50%, and 75% of the way between the full independence with a distribution of diagnoses and full dependence scenarios. We calculated the following FFS adjuster percentages, by the percentage of independence assumed in claims coding errors, as shown in Figure 1.

FIGURE 1: FFS ADJUSTER PERCENTAGES

% OF INDEPENDENCE	FFS ADJUSTER
100% (fully independent) Average diagnoses / beneficiary	1.1%*
100% (fully independent) Actual diagnoses / beneficiary	8.2%
75%	11.6%
50%	14.9%
25%	18.1%
0% (fully dependent)	21.3%

* The scenario using average diagnoses is shown only for reference as a crosswalk from the CMS analysis. Average diagnoses per beneficiary is not a reasonable scenario for calculating a FFS adjuster.

Higher error rates produce similarly larger deviations from actuarial equivalence under the scenario where CMS does not utilize a FFS adjuster in RADV audits. Our simulations of the CMS methodology with varying HCC error rates produce a relatively direct relationship between the error rate and the impact to a FFS adjuster. That is, when the HCC error rates doubled, the deviation from actuarial equivalence also approximately doubled.

COMPUTATIONAL ISSUES

CMS cites the average (mean) error rate at 3% with a median of 2%. CMS does not describe how those estimates were calculated, but based upon the data provided in the Technical Appendix, it appears the error rate it calculated for each HCC was equally weighted without regard to the prevalence of each HCC in the data set.

We utilized the prevalence of HCCs in the 2014 5% Sample and weighted the error rates CMS calculated by HCC to produce an error rate of 2%. This does not impact the results of either the CMS analysis or our analysis. We mention it to identify what may otherwise appear to be an inconsistency in HCC error rates cited in this white paper versus the CMS Technical Appendix.

IS THE SAMPLE SIZE SUFFICIENT?

As a result of CMS's decision to use CERT data, which samples claims rather than beneficiaries for RADV-like reviews of FFS data, it is not possible to definitively determine whether the sample CMS utilized is of sufficient size to be credible to determine the overall HCC error rate. The CMS Technical Appendix asserts statistical calculations to demonstrate the sample is large enough in total, but those statistics require an assumption of independence, which is inappropriate, as previously discussed.

The CMS Technical Appendix does recognize that the error rates they calculate are not credible at the HCC level:

> "One of the principle challenges of using FFS08 for this purpose is that the CERT sample was not designed to produce a representative sample of diagnoses. As a consequence, for many of the diagnoses and by extension, the HCCs, we have an insufficient sample size to develop reliable discrepancy rates at the HCC level. As shown in Table 2a, discrepancy rates ranged from 0-100%. As expected, sample size was an issue for a number of the HCCs. Nearly half of the HCCs had fewer than 28 observations."

As asserted by many MAOs in their criticism of the 2007 RADV audit methodology and demonstrated by CMS in highlighting the widely varying error rates by HCC, the distribution of HCCs in a sample is very important to the results of a RADV audit. CMS has not demonstrated that the sample size utilized in its analysis is large enough to properly calculate a FFS adjuster.

Example demonstrating actuarial equivalence is violated

The theoretical arguments that a FFS adjuster is required in RADV audits are compelling, and we supplement these arguments with concrete examples. The concepts and statistical work required for full calculation of risk scores, calibration, normalization, and RADV audits is extremely complex. Both we and CMS have created simplified examples to highlight the relevant concepts.

CMS EXAMPLE

The CMS developed example is simpler and it was created before the recent proposed rule; however, it does not highlight all of the concepts discussed herein. Appendix A includes this example[22] and clearly demonstrates the need for a FFS adjuster. We acquired this example from the briefs filed in the *UnitedHealthcare Ins. Co. v. Azar* case.

The first table in the CMS example (reproduced in Figure 2 below) shows four beneficiaries, all of whom have diabetes indicated on their claims records. The first three also have diabetes coded in their medical records, while the fourth does not. CMS then lays out an illustrative cost of $4,000 for each beneficiary who has diabetes coded in their medical record. Other conditions and treatments are ignored. This results in a total FFS cost of $12,000 for all four beneficiaries.

[22] The CMS example would be clearer if CMS did not add Beneficiary E in the second slide; however, the result remains the same. This beneficiary increases the initial payment to the plan from the original four beneficiaries but does not change the actual cost to provide care nor does it change the final payment to the plan. In the example as presented, the plan is still underpaid by $3,000 relative to FFS.

Because CMS calibrates and normalizes the HCC model on diagnoses that are on claims, CMS divides the $12,000 of cost by the count of beneficiaries with a diabetes diagnosis on a claim. In this example, there are four beneficiaries, resulting in $12,000 / 4 = $3,000 of cost for each diabetes diagnosis.

FIGURE 2: CMS EXAMPLE, FIRST TABLE

	DIABETES ON CLAIM?	DIABETES IN MEDICAL RECORD?	FFS COST
Beneficiary A	Yes	Yes	$4,000
Beneficiary B	Yes	Yes	$4,000
Beneficiary C	Yes	Yes	$4,000
Beneficiary D	Yes	No	$0
		Total	$12,000
		Diabetes Value for MA Payment	$3,000

The second table in the CMS example (reproduced in Figure 3) demonstrates how an MAO is paid. In this example CMS includes five beneficiaries, all with diabetes coded on claims. The MAO is paid $3,000 each for a total of $15,000. However, Beneficiaries D and E do not have diabetes coded on their medical records. As a result, under a RADV audit, CMS recovers the $6,000 paid to the MAO for Beneficiaries D and E, resulting in a final payment to the MAO of $9,000.

Beneficiaries A through D are identical beneficiaries in the two tables. Under FFS the cost for the four beneficiaries is $12,000,[23] but under the scenario where the MAO undergoes a RADV audit without a FFS adjuster, the MAO is paid $9,000, which is $3,000 less than under FFS.[24] CMS's example clearly demonstrates actuarial equivalence does not exist between FFS and MA when a RADV audit is performed without a FFS adjuster.

FIGURE 3: CMS EXAMPLE, SECOND TABLE

	DIABETES REPORTED BY MA PLAN?	DIABETES IN MEDICAL RECORD?	CMS PAYMENT TO PLAN	PLAN COST	RADV	CMS PAYMENT TO PLAN
Beneficiary A	Yes	Yes	$3,000	$4,000		$3,000
Beneficiary B	Yes	Yes	$3,000	$4,000		$3,000
Beneficiary C	Yes	Yes	$3,000	$4,000		$3,000
Beneficiary D	Yes	No	$3,000	$0	($3,000)	$0
Beneficiary E	Yes	No	$3,000	$0	($3,000)	$0
		Total	$15,000	**$12,000**	($6,000)	**$9,000**

CMS EXAMPLE: EXPANDED

This section expands the prior CMS example with the inclusion of risk scores, normalization, and the calculation of a FFS adjuster to illustrate the normalization effect and the need for a FFS adjuster.

First, for ease of calculations we assume the MAO's bid, that is, the risk-adjusted portion of payments from CMS to the MAO is $10,000 per beneficiary per year. That is, CMS will pay $10,000 to the MAO for a beneficiary with a 1.0 risk score and will pay $11,000 ($10,000 times 1.1) for a beneficiary with a 1.1 risk score.

[23] CMS assumes the plan cost is the same as the FFS cost and that Beneficiaries D and E do not have diabetes, so there is no cost.
[24] In this example, no normalization step is required because total FFS dollar costs are shown; therefore, the $12,000 is already effectively normalized to a 1.0.

We utilize the same four beneficiaries from the CMS example and use the same costs. However, we add a demographic component and assign each beneficiary a different demographic status and cost. Our full example is presented in Appendix B, and we present pieces in tabular format throughout the discussion in this section. Figure 4 shows the four beneficiaries who all have diabetes coded on a claim along with their actual and assumed costs under FFS. We also performed a least squares regression[25] to calibrate our simplified HCC model (which contains three demographic factors and one HCC for diabetes), and the resulting risk score coefficients are shown in Figure 4.

FIGURE 4: EXPANDED DEMOGRAPHICS

TABLE 1

MODEL CALIBRATED AND NORMALIZED WITH UNADJUSTED FFS DIAGNOSES

		FFS COST		
FFS BENEFICIARIES	**ON CLAIM?**	**ACTUAL**	**PREDICTED**	**COEFFICIENT**
Beneficiary 1				
70 yr old			$6,500	0.650
Diabetes	Yes		$3,000	0.300
Subtotal		$9,000	$9,500	0.950
Beneficiary 2				
70 yr old			$6,500	0.650
Diabetes	Yes		$3,000	0.300
Subtotal		$10,000	$9,500	0.950
Beneficiary 3				
75 yr old			$7,000	0.700
Diabetes	Yes		$3,000	0.300
Subtotal		$10,000	$10,000	1.000
Beneficiary 4				
80 yr old Dual			$8,000	0.800
Diabetes	Yes		$3,000	0.300
Subtotal		$11,000	$11,000	1.100
Total		$40,000	$40,000	1.000

For the purposes of this example, we assume these four beneficiaries represent the entire universe of FFS beneficiaries. The total cost for these beneficiaries is $40,000 and we see the model is predicting $40,000 of cost in the "FFS Cost/Predicted" column. Weighting together the coefficients, we see the model produces a 1.000 risk score for the entire FFS population and so is already normalized to a 1.000 risk score, using diagnoses coded on FFS claims (not medical records). The modeling in Figure 4 corresponds to the first model calibration and normalization in the CMS technical analysis, that is, the version where diagnoses are calibrated and normalized on a FFS claims diagnosis basis.

Next, we repeat these steps after reviewing the medical records and finding Beneficiary 4 does not have diabetes documented. We apply least squares regression to recalibrate our simplified HCC model to the medical record diagnoses and calculate the new "FFS Cost/Predicted" and "Coefficients" columns shown in the table in Figure 5.

[25] Due to the simplistic nature of this example, the least squares regression does not produce a unique solution. We used SAS for the regression calculations and seeded the starting values to ensure the particular solution would most resemble the original CMS example we are expanding upon.

FIGURE 5: RECALIBRATED

TABLE 2

MODEL CALIBRATED AND NORMALIZED WITH UNADJUSTED FFS DIAGNOSES

FFS BENEFICIARIES	ON MEDICAL RECORD?	FFS COST ACTUAL	FFS COST PREDICTED	COEFFICIENT
Beneficiary 1				
70 yr old			$5,500	0.550
Diabetes	Yes		$4,000	0.400
Subtotal		$9,000	$9,500	0.950
Beneficiary 2				
70 yr old			$5,500	0.550
Diabetes	Yes		$4,000	0.400
Subtotal		$10,000	$9,500	0.950
Beneficiary 3				
75 yr old			$6,000	0.600
Diabetes	Yes		$4,000	0.400
Subtotal		$10,000	$10,000	1.000
Beneficiary 4				
80 yr old Dual			$11,000	1.100
Diabetes	No		$0	-
Subtotal		$11,000	$11,000	1.100
Total		$40,000	$40,000	1.000

Again, the total actual and predicted FFS cost is $40,000 and our model produces a total risk score of 1.000 when calibrated and normalized using the medical record diagnoses. However, comparing Figures 4 and 5, we observe diabetes has a coefficient of 0.300 in the first scenario and 0.400 in the second scenario. Note the total cost to provide care has not changed and the total risk score for the FFS population is 1.000 in both instances. This second scenario is not performed in the CMS technical analysis, though it should have been because it represents the entire process completed without diagnoses that are not supported on medical records.

The table in Figure 6 illustrates the process that CMS used to develop the revised HCC model in its technical analysis. It shows the risk scores from the model calibrated with the simulated diagnoses documented on medical records in the "Before Normalizing" column. The "On Claim?" column shows a "Yes" where the HCC is applied for a beneficiary, and in this case, shows that the unadjusted claims-based diagnoses are used. Note that the risk scores total to 1.100 for the same four beneficiaries. CMS then applies the normalization step using unadjusted claims-based diagnoses and divides all coefficients by the total risk score for all FFS beneficiaries, which is 1.100. This step is required to ensure the model produces a 1.0 risk score for the FFS population. The resulting new coefficients are in the column, labeled "After Normalizing."

FIGURE 6: CMS PROCESS

TABLE 3

MODEL CALIBRATED WITH ADJUSTED FFS DIAGNOSES BUT NORMALIZED WITH UNADJUSTED DIAGNOSES

FFS BENEFICIARIES	ON CLAIM?	BEFORE NORMALIZING	AFTER NORMALIZING
Beneficiary 1			
70 yr old		0.550	0.500
Diabetes	Yes	0.400	0.364
Subtotal		0.950	0.864
Beneficiary 2			
70 yr old		0.550	0.500
Diabetes	Yes	0.400	0.364
Subtotal		0.950	0.864
Beneficiary 3			
75 yr old		0.600	0.545
Diabetes	Yes	0.400	0.364
Subtotal		1.000	0.909
Beneficiary 4			
80 yr old Dual		1.100	1.000
Diabetes	Yes	0.400	0.364
Subtotal		1.500	1.364
Total		1.100	1.000

* Figures displayed in Figure 6 are rounded to three decimals. Unrounded values are used to produce Figure 7.

Next, we calculate how an MAO would be paid for these identical beneficiaries under four scenarios: (1) no FFS adjuster without a RADV audit, (2) no FFS adjuster with a RADV audit, (3) FFS adjuster without a RADV audit, and (4) FFS adjuster with a RADV audit. The table in Figure 7 shows these four scenarios.

FIGURE 7: HOW MAOS ARE PAID: FOUR SCENARIOS

TABLE 4

	MA PAYMENT WITHOUT FFS ADJUTER			MA PAYMENT WITH FFS ADJUSTER		
FFS BENEFICIARIES	**BEFORE RADV**	**RADV IMPACT**	**AFTER RADV**	**BEFORE RADV**	**RADV IMPACT**	**AFTER RADV**
Beneficiary 1						
70 yr old	$5,000		$5,000	$5,000		$5,000
Diabetes	$3,636		$3,636	$3,636		$3,636
Subtotal	$8,636		$8,636	$8,636		$8,636
Beneficiary 2						
70 yr old	$5,000		$5,000	$5,000		$5,000
Diabetes	$3,636		$3,636	$3,636		$3,636
Subtotal	$8,636		$8,636	$8,636		$8,636
Beneficiary 3						
75 yr old	$5,455		$5,455	$5,455		$5,455
Diabetes	$3,636		$3,636	$3,636		$3,636
Subtotal	$9,091		$9,091	$9,091		$9,091
Beneficiary 4						
80 yr old Dual	$10,000		$10,000	$10,000		$10,000
Diabetes	$3,636	($3,636)	$0	$3,636	($3,636)	$0
Subtotal	$13,636		$10,000	$13,636		$10,000
Total	$40,000		$36,364	$40,000		$36,364
Raw RADV Recovery			$3,636			$3,636
FFS Adjuster			$0			$3,636
Final RADV Recovery			$3,636			$0
Final Payment to MAO	$40,000		$36,364	$40,000		$40,000
Actuarially equivalent?			No			Yes

The payments to the MAO are calculated by multiplying the applicable risk scores or coefficients by the annual MAO bid of $10,000. We utilize the risk scores under the scenario CMS modeled (in the "After Normalizing" column of Figure 6 above), where the model was calibrated with adjusted diagnoses but normalized with unadjusted diagnoses. Unsurprisingly, the two scenarios without a RADV audit produce the same payment as would have been made under FFS, $40,000. However, with a RADV audit, payments to the MAO are reduced to $36,364 because Beneficiary 4 is found to not have diabetes documented in the medical record.

The scenario without a FFS adjuster recovers $3,636 from the MAO, paying the MAO 9% less than would have been paid for identical beneficiaries under FFS, thus violating actuarial equivalence.

To calculate the final payment under the final scenario, with a FFS adjuster, we first must calculate a FFS adjuster. Because we know the risk score under the applicable HCC model for the entire FFS population is 1.100 with claims-based diagnoses and 1.000 with medical records diagnoses, the FFS adjuster is 1.100 divided by 1.000 minus 1, that is, 10%. We calculate the FFS adjuster amount by multiplying the 10% times the payment the RADV audit found to be supported by the medical records, $36,364, and find the FFS adjuster to be $3,636. Finally, the RADV recovery is reduced for the FFS adjuster and the recovery is $0. Under this scenario the MAO is paid $40,000, exactly the same amount as the identical beneficiaries would have cost under FFS. This confirms actuarial equivalence.

For completeness, Appendices C and D repeat the expanded example described here utilizing an HCC model that is calibrated and normalized under the other two scenarios described in this section (Figures 4 and 5). While the size of the FFS adjuster varies, the result is exactly the same. A FFS adjuster is required to maintain actuarial equivalence.

In summary, from these examples it is clear a FFS adjuster is required to maintain actuarial equivalence, as required by statute. The failure to include a FFS adjuster violates actuarial equivalence in every case.

Technical analysis

MODEL AND DATA SELECTION

CMS utilized a model calibration data set of diagnoses from 2004 and claims from 2005 for the FFS portion of the technical analysis. We acquired those data sets in March 2019 when CMS released them. The CMS data set does not include claim level diagnoses that can be mapped to member level demographic and payment data. As a result, certain analyses on the data set cannot be performed. Specifically, when re-calibrating the CMS HCC model using the CMS 2004/2005 data, the actual distribution of the number of claims per HCC cannot be used. To analyze the effect of the CMS simplifying assumption of an average number of diagnoses per beneficiary, we utilized the 2014 and 2015 5% Sample data sets to supplement the FFS portion of our analysis. This data and approach allow us to apply the CMS claim level error rates to claims; and then, to calculate HCC level error rates without assuming an average number of diagnoses per HCC.

As described further in the 'Reproduction of CMS technical approach' section below, we note that our calculation of a FFS adjuster utilizing the CMS data set and the 5% Sample both produced 1.1% under the full independence scenario when using the CMS HCC level error rates and calibrating and normalizing the HCC model to the respective audited data sets.

Similar to CMS, we used version 12 of the CMS HCC model, which was the model in effect for payment years through 2015 (payment years 2014 and 2015 utilized a blend of this model and a newer model.)

We utilized the MA diagnosis data published by CMS in the March 2019 data release to calculate the effect of the various model recalibration scenarios on MA plans.

The particular model or year of data utilized does not impact the conclusion of whether a FFS adjuster is required to maintain actuarial equivalence, though it may impact the magnitude of a FFS adjuster calculated. In the next section, we discuss our reproduction of the CMS results, serving as confirmation that the particular year and version of the model are not material in demonstrating the concepts discussed in this paper.

REPRODUCTION OF CMS TECHNICAL APPROACH

We contacted CMS on several occasions to ensure our interpretation of CMS's analysis was correct. When we contacted CMS directly, CMS cited the Administrative Procedures Act and declined to answer questions and declined to confirm that the text in the Federal Register and the technical backup were correct and as CMS intended. We also asked the same questions on the call CMS hosted to discuss the proposed rule, but the appropriate subject matter experts (SMEs) were not on the phone to answer the questions. CMS also indicated there would be no follow-up call with the SMEs and there would not be time for an FAQ before the end of the comment period. Absent confirmation of our interpretation of the methods CMS utilized, we rely upon the text CMS released, as published, in combination with our reproduction of the methods and the results CMS described.

We reproduced the CMS technical analysis using the CMS data set underlying the technical analysis, as well as the 2015 5% Sample data set (with 2014 diagnoses from the 2014 5% Sample) utilizing the 2013 CMS HCC model. We then applied the recalibrated and renormalized HCC model to the CMS MA HCC data set. In reproducing the CMS methodology, we confirmed that our process also showed that when the CMS HCC model was calibrated with a simulated corrected FFS data set and then normalized with an uncorrected data set, applying the resulting model to MAO beneficiaries does not result in a significant change to MAO risk scores[26].

In June 2019, subsequent to our initial technical analysis, CMS released an Addendum including additional information, additional data, and SAS programs, which further confirmed we correctly understood and reproduced the CMS analysis.

CMS ADDENDUM TO THE FEE-FOR-SERVICE ADJUSTER STUDY AND IPARS

The Addendum included explicit confirmation of technical details we had inferred from prior CMS information releases.

The Addendum also included a mathematical “explanation” of the CMS approach to calculating a calibration bias in the CMS-HCC model in section IV.B., titled “General Expenditure Adjustment to Offset Delete Bias.” The mathematical explanation contains some errors. For example, step 2 defines I_{ji} as the complete matrix of all HCC disease indicators and further that the sumproduct of all coefficients and indicators is equal to the total FFS expenditure (E):

$$\sum_{j=1}^{m}\sum_{i=1}^{p} b_{ji}I_{ji} = \sum_{j=1}^{m} E_j$$

However, the disease indicators do not include demographic variables, which are included in the CMS HCC model and explain a significant portion of expenditures. Further, the use of averages to describe coefficient values in step 5 is inconsistent with Ordinary Least Squares (OLS) because it ignores the difference in weight and frequency of the coefficients and independent variables within the regression model.

If regression concepts were considered rather than average coefficient values, then the removal of a disease indicator for a beneficiary with above average spend for that HCC would decrease, rather than increase (as CMS described in step 6), the coefficient value resulting from OLS.

However, these mathematical problems with the CMS explanation should not be expected to invalidate the overall conclusion that, when the CMS HCC model is calibrated and normalized to produce the total FFS expenditures on separate sets of independent variables, the total always balances to the total FFS expenditure.

By way of this explanation, CMS confirms it asked and answered a question that does not address the need for a FFS adjuster. CMS addressed a question of accuracy in CMS HCC model coefficient calibration but has not calculated a proper FFS adjuster and not addressed actuarial equivalence or the issue of consistently applying the CMS HCC model to the calibration dataset and the payment dataset. We described this issue in the ‘Actuarial equivalence requires a FFS adjuster in RADV’ section and further expound upon it in the ‘CMS technical analysis should not include unsupported FFS diagnoses’ section, above. We illustrate the need for a FFS adjuster in a RADV audit using CMS’s example and an expansion of CMS’s example in the ‘Example demonstrating actuarial equivalence is violated’ section, above. Further, in the next section we discuss one potential adjustment to the CMS approach that could address the question of whether or not a FFS adjuster is required.

Finally, the Addendum repeats the original CMS 50 simulations that measured “audit miscalibration.” CMS completes a new set of 50 simulations, publishing the same results plus an intermediate step that focuses on the ratio of expenses projected by the simulated “corrected” CMS HCC model using “un-perturbed” FFS HCCs to the average actual FFS expenses. CMS refers to this quantity as Inflated Post-

[26] We calculated a mean “audit miscalibration” of 0.002 versus the CMS calculation of 0.001, which we consider to demonstrate successful reproduction of the CMS calculations. Note the calibrated CMS HCC models CMS created in this study do not follow all of the steps CMS uses when creating the final model for actual payment to MA plans and, as such, demonstration of small differences are not sufficient to conclude an actual difference exists.

Audit Risk Scores (IPARS). In the Addendum, CMS calculates IPARS to be 0.9%. The CMS Addendum does not discuss the significance of IPARS; however, a non-zero IPARS demonstrates the need for a FFS adjuster. Further, the CMS IPARS calculation is consistent with our calculation of a payment discrepancy of 1.1% in the next section titled ‘Adjustment of CMS technical approach.’ As demonstrated in the examples and conceptual discussion, above, this difference in risk score and payment results is evidence of the need for a FFS adjuster in RADV audits. If the technical issues with CMS’s estimated HCC error rates were resolved, IPARS would be dramatically larger, emphasizing the critical need for a FFS adjuster.

ADJUSTMENT OF CMS TECHNICAL APPROACH

After confirming we could reproduce the CMS results, we adjusted the normalization process to be completed excluding the simulated unsupported diagnoses. We then applied the new model that was calibrated and normalized on a simulated corrected data set to the MA HCC data and produced MA risk scores, which were, on average, 1% higher than the original model that did not exclude unsupported diagnoses. It is important to note that this 1% effect is certainly material; however, we believe it to be dramatically understated due to the CMS assumptions utilized to create the error rates discussed previously.

We then repeated the analysis, as described throughout this white paper, using a range of HCC error rates that varied by the degree of assumed independence between coding errors from one claim to the next.

To summarize, we completed the following steps to perform an adjusted technical analysis:

1. Filtered diagnoses
    a. Within the CMS data set, we utilized the flag provided by CMS indicating that diagnoses were valid for risk adjustment.
    b. For the 2014 5% Sample data set, we used Encounter Data System (EDS) filtering rules. (We tested for the impact of filtering with Risk Adjustment Processing System [RAPS] rules, found no material difference for the purpose of this study, and elected to use EDS rules for simplicity.)
2. Calibrated the CMS HCC model on unadjusted CMS data / 2014 and 2015 5% Sample data. For the 5% Sample data we utilized the July 2015 cohort of non-ESRD, non-hospice, community population with 12 months of Medicare Part A and Part B enrollment in 2014.
3. Normalized the resulting model to produce a 1.0 risk score for the same total FFS population, again without adjustment to simulate removal of unsupported HCCs.
4. Performed reasonability checks to ensure that the model was reasonably similar to the actual CMS model.
5. Applied the resulting model to the CMS MAO data set to produce a starting point MAO risk score.
6. Set the error rates
    a. For HCC error rate scenarios, set the HCC error rate to be consistent with the particular HCC error rate scenario being processed.
    b. For the claim error rate scenario, set the claim level error rates to the CMS published claim level error rates.
7. Simulation
    a. Simulated adjustments to the filtered CMS data HCCs to produce simulated corrected HCCs.
    b. Simulated claim level adjustments to filtered 2014 diagnoses to produce simulated corrected HCCs.
8. Repeated steps 2 through 5 above, using the simulated corrected HCCs for all steps, including the normalization step.
9. Compared the resulting risk scores for FFS using the original and simulated corrected HCCs under both versions of the HCC models. Under both models using the CMS data, the ratio of the

risk scores using original uncorrected and simulated corrected HCCs was between 1% and 21%, depending upon the assumed HCC error rate. The full independence scenario processed at the claim level on the 5% Sample produced a 12% HCC error rate and an 8% FFS adjuster on the CMS FFS data. These scenarios result in the calculation of a FFS adjuster between 8% and 21% under these assumptions.

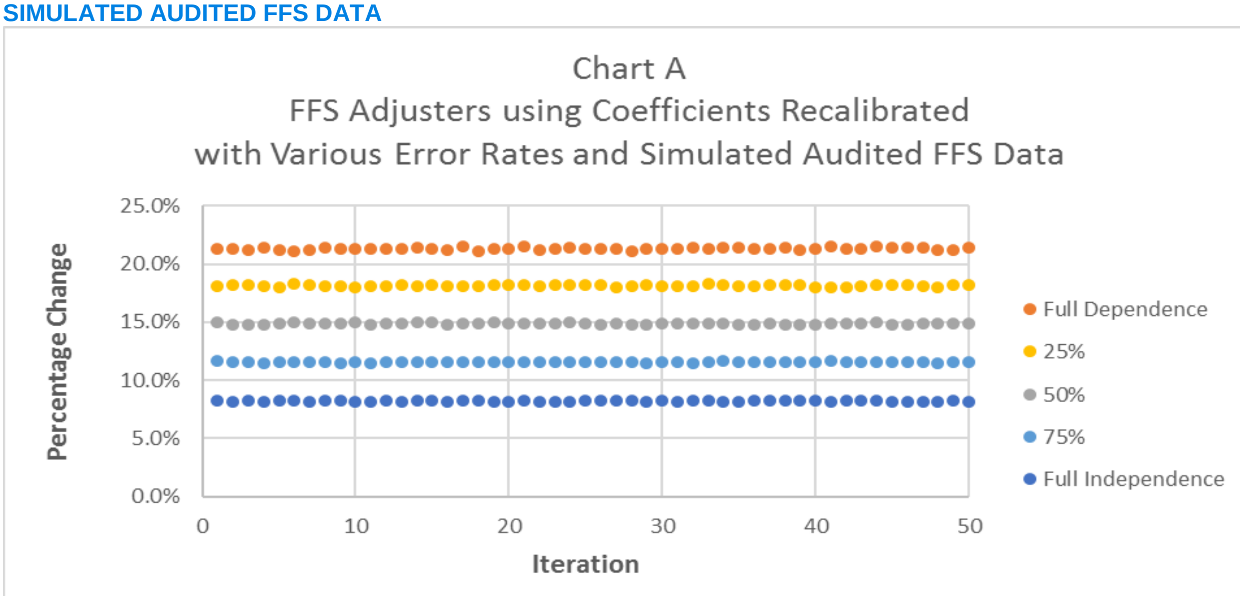
10. Compared the resulting risk scores for MA, based on the CMS MA diagnosis data file, using the original and simulated corrected HCC models. The impact on MA risk scores ranged from 10% to 32%, depending on the level of independence, which is larger than the FFS impact.

Under the midpoint HCC error rate scenario, we performed the simulations and calculated a FFS adjuster of 14.9%, with a range of 8% to 21% for all scenarios (excluding the average claims per HCC scenario). See the chart in Figure 8 for a summary of the key error rate scenarios we calculated. Appendix F, Chart B, contains the same results when calculated on the CMS MA data. We calculate the impact on the MA data as a comparison point to the CMS calculation on MA data included in the technical analysis; however, a FFS adjuster should be calculated on FFS data, not MA data. As discussed earlier in this white paper, properly calculating a FFS adjuster requires performing a credible sampling of FFS beneficiaries and then completing a RADV-type audit on all eligible claims for those beneficiaries. It is not sufficient to calculate error rates for HCCs based upon error rates of individual claims because the degree of independence cannot be known and the results are extremely sensitive to the degree of independence and the distribution of the number of diagnoses per beneficiary. Further study is needed.

50 SIMULATIONS PRODUCE SIMILAR RESULTS

We repeated the adjusted simulation process described above (steps 7 through 10) 50 times for each error rate scenario, as CMS did with its version of the analysis (but using a single error rate). We observed minimal variations in the resulting value of the FFS adjuster within each error rate scenario. Figure 8 shows the consistency of the FFS adjuster results across error rate scenario simulations. Appendix F includes additional exhibits showing consistency of the impact on MA risk scores across simulations and highlighting selected key distributional statistics.

FIGURE 8: FFS ADJUSTERS USING COEFFICIENTS RECALIBRATED WITH VARIOUS ERROR RATES AND SIMULATED AUDITED FFS DATA



Under the midpoint error rate scenario and based upon 50 iterations, the FFS adjuster is between 14.85% and 14.90%, with a 99% level of confidence.

Context around the CMS HCC risk model

As set out in statute, the CMS HCC model is intended to adjust payment amounts made to MAOs by beneficiary health status. The HHS Secretary has broad authority to add or remove adjustment factors if such changes will improve the determination of actuarial equivalence, which further highlights the emphasis on actuarial equivalence from Congress. Beyond requiring a risk adjustment model and actuarial equivalence, the statute goes on to require an adjustment for the coding pattern difference between FFS and MA. Understanding appropriate creation and application of the risk score model also requires an understanding of the background, procedures, and adjustments surrounding the implementation of the risk score model.

RISK MODEL DESIGN

Several considerations should go into designing a risk score model. In the case of the CMS HCC model, a strong model would compensate MAOs for the health status of the beneficiaries they enroll without creating an incentive to enroll certain types of beneficiaries over others. A strong risk adjustment model could be based upon diagnoses from medical records, because these diagnoses most closely reflect the actual conditions beneficiaries are treated for. Given that this is impractical from an administrative cost perspective, CMS needed data to serve as a proxy for medical record diagnoses. To fill this void, CMS designed the CMS HCC model utilizing diagnoses from claims data. While providers have not historically had a strong incentive to accurately report diagnoses in claims data (with the exception of inpatient claims), the claims-based diagnoses are a reasonable proxy for medical record diagnoses in the context of establishing the disease burden of an individual beneficiary.

Predictive models and risk score models are often measured based upon how well they predict results for individual beneficiaries. However, as CMS points out in the “Weak Statistical Foundations” section of the Technical Appendix, referenced in the proposed rule, MAOs are paid to provide care for an entire population of beneficiaries. It is important to pay MAOs accurately for the entire population of beneficiaries, but is less important to pay MAOs correctly for each individual beneficiary and HCC. As CMS lays out with mathematical formulas, if the actual cost of providing care for a beneficiary with a particular HCC varies significantly, the quality of the risk score model, in the context of paying MAOs, is not reduced as long as variation from the average cost of providing care is not biased. That is, the quality of the model is not reduced if the cost to provide care above the average and below the average for an HCC are approximately equivalent. The commentary in the “Weak Statistical Foundations” section may be important in establishing a good risk score payment model, but it has no relevance for actuarial equivalence or a FFS adjuster.

CMS goes on to discuss, in the Technical Appendix, a concept it refers to as “Calibration Error Correction Limited to Recoveries is Economically Problematic.” The arguments put forth focus upon the concept that there may be calibration errors in the CMS HCC model. While calibration errors may impact the relative values of one HCC against another, they have little bearing on total payments as a result of the CMS step that normalizes the CMS HCC model to a 1.0 risk score for the FFS population. While minimizing calibration error may be important to developing a risk model, this topic is also not relevant to actuarial equivalence or a FFS adjuster.

MODEL IMPLEMENTATION

CMS’s Risk Adjustment Participant Guides focus upon rules and guidelines for plans to filter claims data and submit the diagnoses attached to such claims through the Risk Adjustment Processing System (RAPS). That is, CMS publishes the rules by which plans must abide when submitting claims-based diagnosis data. RADV audits, however, do not primarily measure how well a plan complies with the filtering and submission process set forth by CMS. Rather, the RADV audit compares the claims-based diagnoses to the diagnoses on the medical charts and cites the differences as errors made by the plan. Therefore, RADV audit procedures primarily measure how well claims-based diagnoses approximate medical chart diagnoses.

The RADV audit process primarily measures the bias of the diagnosis proxy, that is, the difference between claims-based diagnoses and medical record diagnoses. Such a bias exists on both the FFS data and the MA data. Title 42 U.S. Code § 1395w–23(a)(1)(C)(i) requires the risk model to "ensure actuarial equivalence" between FFS and MA. Removing the bias from either side without removing it from the other compromises the risk adjustment model by violating actuarial equivalence, and therefore statute. If the bias is removed from the MAO side but not the FFS side, one solution to maintain actuarial equivalence is to apply a FFS adjuster in the implementation of the RADV audit. The addition of a FFS adjuster is akin to adjusting the CMS HCC model to be on a medical record diagnosis basis, consistent with the methodology of the RADV audit for MA diagnosis support.

OTHER ADJUSTMENT FACTORS

CMS implements other adjustments surrounding the risk score model and its implementation. A few of these adjustments are discussed here for completeness.

FFS normalization: Provider coding patterns change over time and the FFS Medicare population changes over time. Because the data required to create and calibrate an HCC model is several years old, CMS must project both changes in the FFS population and FFS provider coding practices in an attempt to maintain a 1.0 risk score for future years. The FFS normalization factor is the CMS projected estimate of what the risk score of the FFS population will be in a future payment year. All risk scores are then divided by this factor. This concept is very similar to the normalization step discussed throughout much of this white paper.

Medicare Secondary Payer (MSP) adjustment: Certain beneficiaries have medical insurance aside from Medicare. For those beneficiaries who have other coverage that pays primary to Medicare, CMS estimates a reduction to Medicare's expense for those beneficiaries. This reduction is generally over 80% and is applied in the MA bid process as a reduction to risk scores.

MA coding pattern adjustment: The MA coding pattern adjustment is intended to capture any difference between how FFS and MA beneficiary diagnoses are coded. The difference between claims-based diagnoses and medical record-based diagnoses may be different between FFS and MA. To the extent they are different, that difference between the documentation error rates may already be included in the MA coding pattern adjustment. Further study would be required to separate the impact of a true coding pattern adjustment from a difference in the way claims-based diagnoses and medical record-based diagnoses vary between FFS and MA.

Other considerations for calculating FFS adjusters

This white paper is focused primarily upon overall actuarial equivalence between FFS and MA and properly calculating error rates (generally the difference between claims and medical record diagnoses.) There are other considerations for calculating a final FFS adjuster, or simply performing a more precise analysis regarding the need for one.

The CMS technical analysis used a variety of data from a variety of time periods. The ICD-9 to HCC mapping was from a single time period and so may not be consistent with portions of the underlying data. As ICD-9 codes do change over time, and CMS updates the mapping over time, the applicable year's model should be used.

The CMS technical analysis uses random numbers to simulate unsupported diagnoses. However, the CMS HCC model is built on a causal relationship between diagnoses and claims. A proper analysis of accuracy in model calibration must use actual coding errors to maintain the assumed causal relationship of the HCC model, not randomized changes. Stated more technically, OLS (Ordinary Least Squares) regression measures correlation between dependent and independent variables. As such, modifying the independent variables in a random fashion compromises correlation and any conclusions drawn from OLS.

Further, error rates should be expected to change as CMS updates the HCC models and the mappings within them. For example, the 2014 HCC model included a clinical revision that was at least partially intended to address some of the coding differences present in MA versus FFS and this should be expected to impact the error rates. Provider coding practices change over time and should have an effect on error rates. The advent of ICD-10 during the fourth quarter of 2015 and the ever-increasing penetration of electronic medical records should also be expected to change the error rates over time.

As CMS considers different time periods, the error rates should be revisited and recalculated frequently to reflect the applicable time period's models, error rates, and coding practices.

Additional statistical background

Least squares regression approaches are a category of statistical methodologies intended to minimize the sum of the squares of the residuals. The residuals are the difference between the observed data used for calibrating the model and the amount predicted for that data point. These residuals are raised to the second power (squared) and then added across all observed data points. The residuals can be thought of as amounts that the calibrated model does not predict. The goal of least squares regression is to minimize the square of the residuals (error terms).

OLS methodologies weight each data point equally, while weighted least squares applies a weight to each data point, for example the amount of claims or the number of months a beneficiary is enrolled for in the projection year of the CMS HCC model.

Conclusion

CMS currently calibrates and normalizes the CMS HCC model on FFS data that is based upon diagnoses from claims records. Because RADV audits utilize medical records and a different coding standard, RADV findings must be adjusted by the difference between those coding standards within FFS, that is, a FFS adjuster. Failure to make an adjustment, such as a FFS adjuster in the context of RADV audits and the current risk adjustment system, violates actuarial equivalence, and actuarial equivalence is required by federal law.

The CMS technical analysis accompanying the proposed rule did not state CMS calculated a FFS adjuster and did not appropriately calculate a FFS adjuster in the context of RADV audits. Instead, it measured a calibration bias of a CMS HCC model, which does not answer the question of whether or not a FFS adjuster is required. At a minimum, an analysis of a FFS adjuster must exclude unsupported diagnoses from all steps of the calibration and normalization process. Since the CMS analysis does not exclude unsupported diagnoses from the normalization process, it cannot be used to support the removal of a FFS adjuster.

Estimation of a FFS adjuster should be based upon data and models that are consistent with the data that will undergo a RADV audit. Further, FFS beneficiaries should be sampled and, at a minimum, all claims containing diagnoses mapping to HCCs should be audited. Error rates should then be calculated while considering the beneficiary as a whole and including diagnoses for which the provider does not provide documentation, which is how beneficiaries are treated for payment and how beneficiaries are evaluated for HCCs.

Appendix A: CMS documents from Docket 44. United v Price No. 1:16-cv-00157-RMC

Why does FFS Diagnosis Error Matter?

	Diabetes on Claim?	Diabetes in medical record?	FFS Cost
Beneficiary A	Yes	Yes	$4,000
Beneficiary B	Yes	Yes	$4,000
Beneficiary C	Yes	Yes	$4,000
Beneficiary D	Yes	No	$0
		Total	$12,000
		Diabetes Value for MA Payment	$3,000

	Diabetes reported by MA plan?	Diabetes in medical record?	CMS Payment to Plan	Plan Cost	RADV	CMS Payment to Plan
Beneficiary A	Yes	Yes	$3,000	$4,000		$3,000
Beneficiary B	Yes	Yes	$3,000	$4,000		$3,000
Beneficiary C	Yes	Yes	$3,000	$4,000		$3,000
Beneficiary D	Yes	No	$3,000	$0	($3,000)	$0
Beneficiary E	Yes	No	$3,000	$0	($3,000)	$0
		Total	$15,000	**$12,000**	($6,000)	**$9,000**

Appendix B: Full expanded example of calibration and normalization of HCC model: Calibrated with adjusted diagnoses and normalized with unadjusted diagnoses: CMS proposed rule technical analysis approach

	MODEL CALIBRATED AND NORMALIZED WITH UNADJUSTED FFS DIAGNOSES				MODEL CALIBRATED AND NORMALIZED WITH ADJUSTED FFS DIAGNOSES			
FFS BENEFICIARIES	ON CLAIM?	ACTUAL FFS COST	PREDICTED FFS COST	COEFFICIENT	ON MEDICAL RECORD?	ACTUAL FFS COST	PREDICTED FFS COST	COEFFICIENT
Beneficiary 1								
70 yr old			$6,500	0.650			$5,500	0.550
Diabetes	Yes		$3,000	0.300	Yes		$4,000	0.400
Subtotal		$9,000	$9,500	0.950		$9,000	$9,500	0.950
Beneficiary 2								
70 yr old			$6,500	0.650			$5,500	0.550
Diabetes	Yes		$3,000	0.300	Yes		$4,000	0.400
Subtotal		$10,000	$9,500	0.950		$10,000	$9,500	0.950
Beneficiary 3								
75 yr old			$7,000	0.700			$6,000	0.600
Diabetes	Yes		$3,000	0.300	Yes		$4,000	0.400
Subtotal		$10,000	$10,000	1.000		$10,000	$10,000	1.000
Beneficiary 4								
80 yr old Dual			$8,000	0.800			$11,000	1.100
Diabetes	Yes		$3,000	0.300	No		$0	-
Subtotal		$11,000	$11,000	1.100		$11,000	$11,000	1.100
Total		$40,000	$40,000	1.000		$40,000	$40,000	1.000

		MODEL CALIBRATED WITH ADJUSTED FFS DIAGNOSES BUT NORMALIZED WITH UNADJUSTED DIAGNOSES		MA PAYMENT WITHOUT FFS ADJUSTER		MA PAYMENT WITH FFS ADJUSTER	
FFS BENEFICIARIES	**ON CLAIM?**	**BEFORE NORMALIZING**	**AFTER NORMALIZING**	**BEFORE RADV**	**AFTER RADV**	**BEFORE RADV**	**AFTER RADV**
Beneficiary 1							
70 yr old		0.550	0.500	$5,000	$5,000	$5,000	$5,000
Diabetes	Yes	0.400	0.364	$3,636	$3,636	$3,636	$3,636
Subtotal		0.950	0.864	$8,636	$8,636	$8,636	$8,636
Beneficiary 2							
70 yr old		0.550	0.500	$5,000	$5,000	$5,000	$5,000
Diabetes	Yes	0.400	0.364	$3,636	$3,636	$3,636	$3,636
Subtotal		0.950	0.864	$8,636	$8,636	$8,636	$8,636
Beneficiary 3							
75 yr old		0.600	0.545	$5,455	$5,455	$5,455	$5,455
Diabetes	Yes	0.400	0.364	$3,636	$3,636	$3,636	$3,636
Subtotal		1.000	0.909	$9,091	$9,091	$9,091	$9,091
Beneficiary 4							
80 yr old Dual		1.100	1.000	$10,000	$10,000	$10,000	$10,000
Diabetes	Yes	0.400	0.364	$3,636	$0	$3,636	$0
Subtotal		1.500	1.364	$13,636	$10,000	$13,636	$10,000
Total		1.100	1.000	$40,000	$36,364	$40,000	$36,364
Raw RADV Recovery					$3,636		$3,636
FFS Adjuster					$0		$3,636
Final RADV Recovery					$3,636		$0
Final Payment to MAO				$40,000	$36,364	$40,000	$40,000
Actuarially equivalent?*				Yes	No	Yes	Yes

* When the CMS HCC model is normalized with unadjusted diagnoses, actuarial equivalence is maintained at initial payment and under a RADV audit with a FFS adjuster, not with a RADV audit without a FFS adjuster.

Appendix C: Full expanded example of calibration and normalization of HCC model: Calibrated and normalized with adjusted diagnoses

	MODEL CALIBRATED AND NORMALIZED WITH ADJUSTED FFS DIAGNOSES		MA PAYMENT WITHOUT FFS ADJUSTER	
FFS BENEFICIARIES	**ON CLAIM?**	**AFTER NORMALIZING**	**BEFORE RADV**	**AFTER RADV***
Beneficiary 1				
70 yr old		0.550	$5,500	$5,500
Diabetes	Yes	0.400	$4,000	$4,000
Subtotal		0.950	$9,500	$9,500
Beneficiary 2				
70 yr old		0.550	$5,500	$5,500
Diabetes	Yes	0.400	$4,000	$4,000
Subtotal		0.950	$9,500	$9,500
Beneficiary 3				
75 yr old		0.600	$6,000	$6,000
Diabetes	Yes	0.400	$4,000	$4,000
Subtotal		1.000	$10,000	$10,000
Beneficiary 4				
80 yr old Dual		1.100	$11,000	$11,000
Diabetes	Yes	0.400	$4,000	$0
Subtotal		1.500	$15,000	$11,000
Total		1.100	$44,000	$40,000
Raw RADV Recovery				$4,000
FFS Adjuster				$0
Final RADV Recovery				$4,000
Final Payment to MAO			$40,000	$40,000
Actuarially equivalent?*			No	Yes

* When the CMS HCC model is normalized with adjusted diagnoses, a FFS adjuster is not required and actuarial equivalence is achieved only after a RADV audit.

Appendix D: Full expanded example of calibration and normalization of HCC model: Calibrated and normalized with unadjusted diagnoses, status quo before the proposed rule

	MODEL CALIBRATED AND NORMALIZED WITH UNADJUSTED FFS DIAGNOSES		MA PAYMENT WITHOUT FFS ADJUSTER*		MA PAYMENT WITH FFS ADJUSTER*	
FFS BENEFICIARIES	ON CLAIM?	AFTER NORMALIZING	BEFORE RADV	AFTER RADV	BEFORE RADV	AFTER RADV
Beneficiary 1						
70 yr old		0.650	$6,500	$6,500	$6,500	$6,500
Diabetes	Yes	0.300	$3,000	$3,000	$3,000	$3,000
Subtotal		0.950	$9,500	$9,500	$9,500	$9,500
Beneficiary 2						
70 yr old		0.650	$6,500	$6,500	$6,500	$6,500
Diabetes	Yes	0.300	$3,000	$3,000	$3,000	$3,000
Subtotal		0.950	$9,500	$9,500	$9,500	$9,500
Beneficiary 3						
75 yr old		0.700	$7,000	$7,000	$7,000	$7,000
Diabetes	Yes	0.300	$3,000	$3,000	$3,000	$3,000
Subtotal		1.000	$10,000	$10,000	$10,000	$10,000
Beneficiary 4						
80 yr old Dual		0.800	$8,000	$8,000	$8,000	$8,000
Diabetes	Yes	0.300	$3,000	$0	$3,000	$0
Subtotal		1.100	$11,000	$8,000	$11,000	$8,000
Total		1.000	$40,000	$37,000	$40,000	$37,000
Raw RADV Recovery				$3,000		$3,000
FFS Adjuster				$0		$3,000
Final RADV Recovery				$3,000		$0
Final Payment to MAO			$40,000	$37,000	$40,000	$40,000
Actuarially equivalent?			Yes	No	Yes	Yes

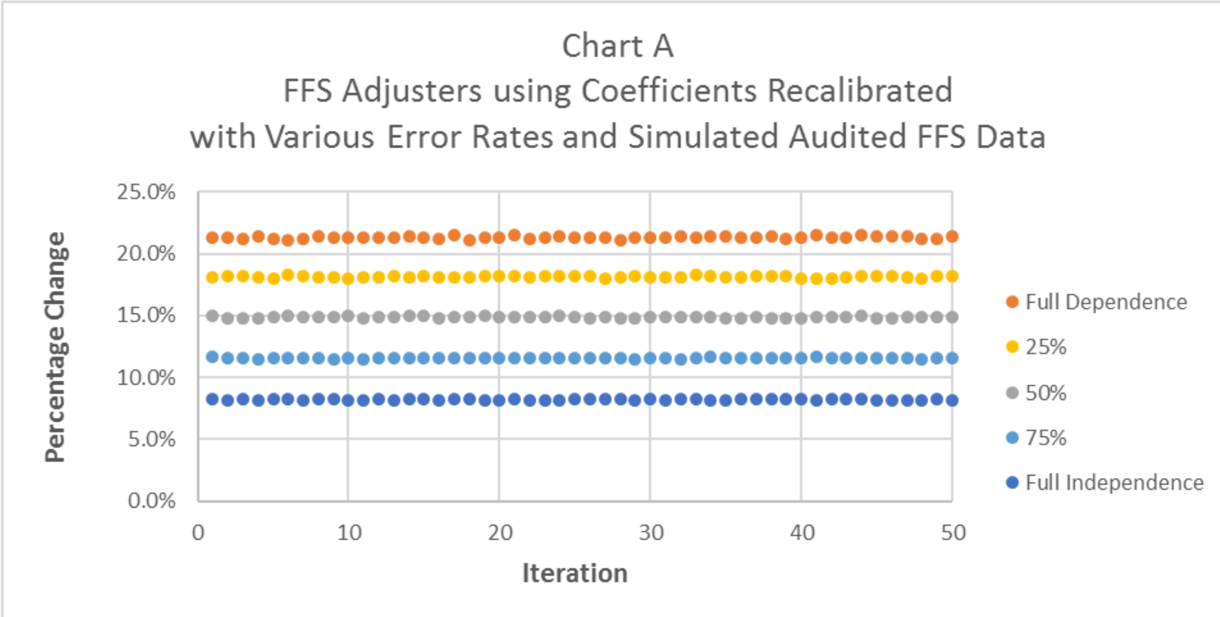
* When the CMS HCC model is normalized with unadjusted diagnoses, actuarial equivalence is maintained at initial payment and under a RADV audit with a FFS adjuster, not with a RADV audit without a FFS adjuster.

Appendix E: CMS documents from Docket 44. United v Price No. 1:16-cv-00157-RMC

Model Calibration Factor.

- The first issue is the extrapolation methodology that we're going to use in RADV.
- The approach that we laid out in our December guidance was pretty straightforward and we are not recommending making any significant changes -- with one possible exception.
- Plans have raised the concern that we are holding them to a standard of perfection for diagnosis coding but that physician claims in FFS Medicare often include diagnoses that aren't supported in the medical record.
- So they argue that we have two different documentation standards – one for MA and one for FFS.
- And this wouldn't matter except that we use FFS claims data to develop our risk adjustors for Medicare Advantage.
- In risk adjustment model, we are estimating the average relative cost of any given condition given the people who are reported to have it.
- So when we estimate the relative cost of any given condition, we use diagnosis and cost data from FFS Medicare.
- So implicit in all of the adjustments we make to plans payments to account for the relative risk of their populations, are the factors that we developed using FFS data.
- If we include diagnoses for beneficiaries who don't actually have the disease, or for whom the medical record documentation is not clear, this tends to reduce the estimated average cost of various conditions and therefore our risk adjustment factors.
- [redacted] (b)(5)
- So plans argue that we are paying them as if they are getting beneficiaries who look like FFS rather than the higher average cost of the beneficiaries we are allowing to be claimed in MA under the RADV audits.
- The address this issue, we are proposing to develop a model calibration factor that estimates how much higher the plan's payment would be if our risk adjustment model had been built using perfect data.
- This factor would reduce the estimated RADV overpayments due from the plan.
- We think this approach makes sense and from a technical point of view is the right thing to do.
- It also will help bring the overpayments into a range that is more realistic for plans to be able to accommodate.

Appendix F: Statistical results from 50 simulations



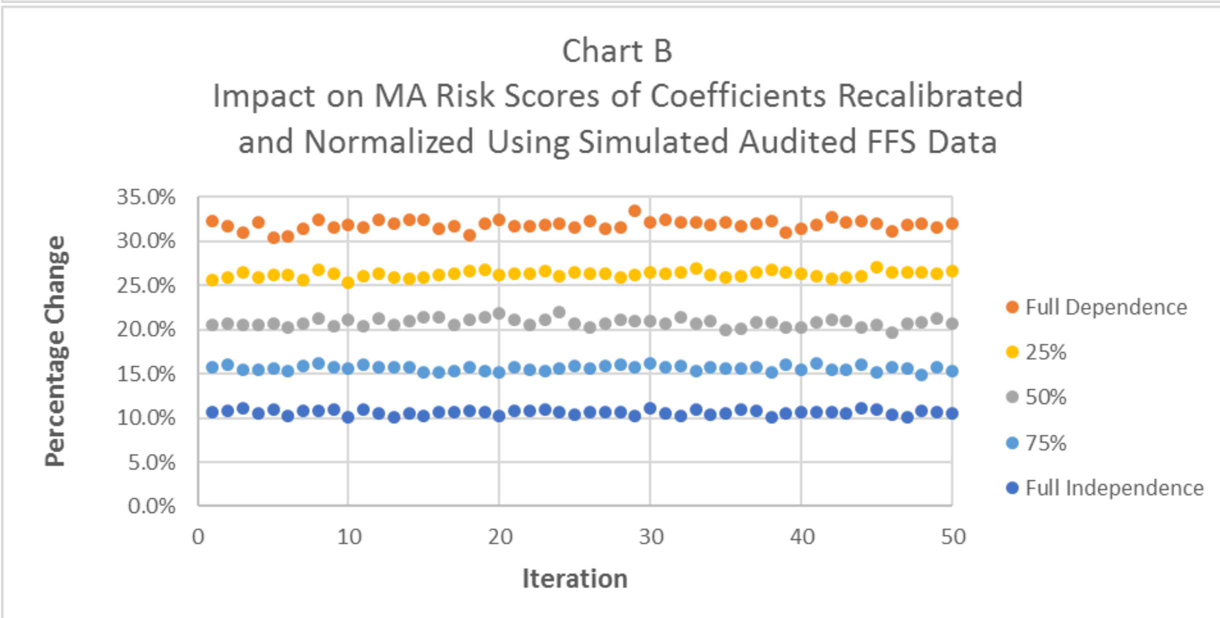


Table 5: FFS Distributional Statistics

	Degree of Independence				
	0%	25%	50%	75%	100%
Mean FFS adjuster	21.3%	18.1%	14.9%	11.6%	8.2%
Median FFS adjuster	21.3%	18.1%	14.9%	11.6%	8.2%
Minimum FFS adjuster	21.1%	18.0%	14.7%	11.5%	8.1%
Maximum FFS adjuster	21.5%	18.3%	15.0%	11.7%	8.3%
25th Percentile	21.3%	18.1%	14.8%	11.5%	8.2%
75th Percentile	21.4%	18.2%	14.9%	11.6%	8.2%
Sample Standard Deviation	0.09%	0.07%	0.06%	0.04%	0.04%
Lower 99% Confidence Bound	21.3%	18.1%	14.9%	11.5%	8.2%
Upper 99% Confidence Bound	21.3%	18.2%	14.9%	11.6%	8.2%