August 27, 2019

Ms. Seema Verma, Administrator
Centers for Medicare & Medicaid Services
Department of Health and Human Services
Hubert H. Humphrey Building, Room 445-G
200 Independence Avenue, S.W.
Washington, D.C. 20201

**RE: 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 ("Proposed Rule")**

Dear Administrator Verma:

America's Health Insurance Plans (AHIP) appreciates the opportunity to comment on the Risk
Adjustment Data Validation (RADV) proposal put forward by the Center for Program Integrity (CPI)
and included in the Proposed Rule. AHIP is the national association whose members provide
coverage for health care and related services for millions of Americans. Through these offerings,
including Medicare Advantage (MA) plans, we improve and protect the health and financial security
of consumers, families, businesses, communities, and the nation. We are committed to market-based
solutions and public-private partnerships that improve affordability, value, access, and well-being for
consumers. The MA program is critical to achieving national policy goals for improved health care,
and we share your strong commitment to delivering better health outcomes, value, and satisfaction to
Medicare beneficiaries.

Proposed Rule Is Fatally Flawed and Should Be Withdrawn
The MA program and its payment structure are designed to encourage MA plans to maximize the
efficient provision of high-quality health care treatments and services to Medicare beneficiaries.
They are also designed to ensure that MA plans have the resources needed to provide high-quality
benefits and coordinated care to seniors and people with disabilities.

As part of this structure, Congress directed CMS to adjust payments to each plan to account for the
health status of its population.[1] CMS carries out this mandate through a risk adjustment model that is
based on data from traditional Medicare. Congress also required CMS to ensure the risk adjustment
model achieves actuarial equivalence between MA and traditional Medicare.[2] This actuarial
equivalence requirement is fundamental to the proper functioning of the MA risk adjustment model,
and therefore is a core component of a stable MA program.

---

[1] 42 USC 1395w-23(a)(3).

[2] 42 USC 1395w-23(a)(1)(C)(i) – "…the Secretary shall adjust the payment…for such risk factors as age, disability
status, gender, institutional status, and such other factors as the Secretary determines to be appropriate, including
adjustment for health status under paragraph (3), *so as to ensure actuarial equivalence*." (emphasis added)

As we note in our comments below, we believe the Proposed Rule violates this critical actuarial equivalence requirement. First and foremost, it fails to include a fee-for-service (FFS) adjuster to ensure actuarial equivalence between payments to MA plans and payments under the traditional Medicare program.[3] The rule also suffers from other serious substantive and procedural defects. Given our very strong legal and policy objections, we urge CMS to withdraw the RADV proposal.

Proposed Rule Undermines MA and Confidence in Government as Fair Business Partner
We appreciate CMS' decision to provide data and additional methodological explanations regarding its technical study relating to the proposal in response to requests from stakeholders and the agency's decision to extend the comment deadline on the RADV proposal, accordingly. However, these data and explanations do not solve or mitigate our serious concerns with the RADV proposal.

Health insurance providers are accountable to the consumers they serve as well as the taxpayers who fund the MA program. Since 2010, MA plan sponsors have requested that CMS engage in a dialogue to develop a fair and appropriate oversight process. Rather than engage with its private-sector partners, CPI released a proposal that is neither fair nor appropriate. It would reverse the well-established and long-held principle that a FFS adjuster is necessary to meet statutory and actuarial requirements.[4] The Proposed Rule would also permit actions that exceed the agency's legal authority, including collecting contract-wide payment amounts and retroactively changing rules going back almost a decade (to 2011).

If CMS were to finalize the RADV provisions in the Proposed Rule, it would undermine stakeholder confidence in the agency's willingness to comply with the law and to act as a fair partner with the private sector. Private-sector partners must be able to rely on the government's word and know that the government will adhere to its commitments, whether stemming from statute or otherwise. A lack of trust injects significant uncertainty and risk into the system, undermines how the free market and public programs work together, and fundamentally weakens the integrity of the MA program. As a result, seniors and hardworking taxpayers might see higher costs, reduced benefits, and fewer MA plan options.

There is a better way. We ask that CMS withdraw these provisions and work with us on real solutions that are fair, accurate, and legally permissible.

Growing Value and Attractiveness of MA over Traditional Medicare
MA plans deliver better care and better value through innovative, patient-centered programs that improve quality and reduce costs. In the past decade, enrollment in MA has nearly doubled. More

---

[3] The FFS adjuster accounts for the fact that the documentation standard used in RADV – that claims must be submitted absolutely free of diagnosis coding errors – is different from the documentation standard used to calculate the risk adjustment model, which includes unsubstantiated FFS claims data with diagnosis coding errors.

[4] Centers for Medicare & Medicaid Services, Notice of Final Payment Error Calculation Methodology for Part C Medicare Advantage Risk Adjustment Data Validation Contract-Level Audits, February 24, 2012. "…to determine the final payment recovery amount, CMS will apply a Fee-for-Service Adjuster (FFS Adjuster) amount as an offset to the preliminary recovery amount…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)." Available at: 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.

than 22 million Americans – over one-third of all Medicare beneficiaries – have chosen to enroll in MA plans. These plans provide financial security by limiting out-of-pocket costs, offering integrated drug coverage, and providing a rich array of benefits not available in traditional Medicare, including dental, vision, hearing, and other supplemental benefits. Enrollees in MA plans are highly satisfied with the MA program.[5] MA plans provide these benefits at the same cost as traditional Medicare.[6] And in areas of the country where MA is popular, additional enrollment leads to slower traditional Medicare spending growth as providers employ MA practice patterns and care guidelines for their remaining traditional Medicare patients.[7]

Summary of Key RADV Concerns
We have attached comments with significantly more detail about our legal and policy concerns with the RADV changes in the Proposed Rule. They include the following:

- **A FFS Adjuster is required to ensure actuarial equivalence.** We have attached an analysis from the actuarial firm Milliman, based on CMS' data and methodologies as presented in the CMS technical study included with the Proposed Rule and subsequent data releases, that shows a FFS adjuster, or other similar adjustment, is necessary to ensure actuarial equivalence between payments to MA plans and payments under the FFS program. This adjustment is required due to the different documentation standards for the determination of diagnoses under MA and traditional Medicare. Milliman makes adjustments to address errors in CMS' methodology to find that a FFS adjuster would be both positive and material, and concludes that the CMS technical analysis "cannot appropriately be used to conclude a FFS adjuster is not required." Another study highlighted in our comments, and two recent federal district court decisions, also conclude that a FFS adjuster is required. In addition, the CMS technical study and addendum fail to address the key issue of actuarial equivalence in the context of RADV audits, contain multiple flaws and questionable assumptions, and in short, appear to have been designed to minimize error rates, enabling CMS to arrive at the conclusion that a FFS adjuster is not warranted. Further, CMS' claim that a FFS adjuster would create inequity among plans is neither credible, reasonable, nor consistent with a recent court decision which confirms that the statute requires CMS adjust payments to ensure actuarial equivalence. We believe strongly that CMS is required to implement a FFS adjuster in payment recovery activities, and that such an adjustment is not only necessary to achieve actuarial equivalence but is equitable for both audited and unaudited plans in that context.

- **CPI has no legal authority for extrapolation and, even if it did, is proposing to use a flawed methodology.** In RADV audits, CMS reviews medical records from a sample of beneficiaries. The Proposed Rule would provide for "extrapolation" – i.e., CMS would use the sample to calculate a contract-wide error rate and recover payments accordingly. The Social Security Act (SSA) provides authority to extrapolate, but only for Medicare

---

[5] Morning Consult National Poll. November 28-29, 2018. In this poll, 90 percent of MA members reported satisfaction with their health care coverage and preventive services, and 84 percent reported satisfaction with their prescription drug coverage.
[6] Medicare Payment Advisory Commission. Report to the Congress: Medicare payment policy. March 2019. For 2019, MA plan payments are equivalent to traditional Medicare costs.
[7] Johnson, Garret, Figuero, Jose F., Zhou, Xiner, et al. Recent growth in Medicare Advantage enrollment associated with decreased fee-for-service spending in certain US counties. *Health Affairs* 35(9): 1707-1715. September 2016.

contractors auditing providers under Parts A and B, and only in limited circumstances.[8] The Proposed Rule provides no legal justification for extrapolation. CPI simply asserts it has such authority.[9] Further, based on the findings of a study by Wakely (attached) that identified several significant areas of concern, we believe the extrapolation methodology CMS published in 2012 raises serious policy concerns because it will produce arbitrary results.[10]

- **Retroactivity is prohibited by federal law and is unnecessary and unjustified.** CMS proposes to grant itself authority to extrapolate audits for plan years 2011 and forward. The SSA clearly prohibits retroactive rules absent a statutory requirement, significant public safety concern, or other critical need[11] – none of which are present here. In addition, retroactivity poses major operational barriers for plans and providers. For example, CPI has recently begun conducting RADV audits for 2014 which review services rendered in 2013. This long passage of time could make it extremely difficult for plans to obtain medical records with respect to providers who, for example, are deceased, closed their practices, changed to new recordkeeping systems, etc.

- **Implementation of the RADV audit methodology would violate rulemaking requirements.** The methodology used by CPI for 2011, 2012, and 2013 audits was never subject to prior notice and comment rulemaking. And while CPI is now soliciting comments, the Proposed Rule indicates that CPI can implement changes solely through Health Plan Management System (HPMS) notices. CPI also has begun moving forward with audits for 2014, using a new methodology, without allowing a comment opportunity or broadly providing any details to the public on the methodology it is using. These actions are inconsistent with the SSA requirements for notice-and-comment rulemaking as indicated by the U.S. Supreme Court in its recent ruling in *Azar v. Allina Health Services.*[12]

RADV Recommendations
Based upon the legal and policy reasons above, we urge CMS to take the following steps:

- **Withdraw the RADV proposal.** The RADV provisions in the Proposed Rule should not be finalized. The provisions should be withdrawn in their entirety so together we can develop a collaborative and constructive solution.

---

[8] 42 U.S.C. 1395ddd.

[9] See, e.g., 83 Fed. Reg. 54982, 54984 (Nov 1, 2018), where CMS asserts that extrapolation is "based on longstanding case law and best practices from HHS and other federal agencies" but provides no citations to or analysis of this authority.

[10] Murray, T., Morgan, E., Sauter, M. Medicare RADV: Review of CMS sampling and extrapolation methodology. Wakely Consulting Group. July 2018. Available at: https://www.ahip.org/wp-content/uploads/2018/07/Wakely-Medicare-RADV-Report-2018.07.pdf.

[11] 42 USC 1395hh(e)(1)(A) – "A substantive change in regulations, manual instructions, interpretative rules, statements of policy, or guidelines of general applicability under this subchapter shall not be applied (by extrapolation or otherwise) retroactively to items and services furnished before the effective date of the change, unless the Secretary determines that- (i) such retroactive application is necessary to comply with statutory requirements; or (ii) failure to apply the change retroactively would be contrary to the public interest."

[12] 139 S.Ct. 1804 (2019).

- **Affirm that regulatory changes cannot be applied retroactively.** CMS should follow the clear directive in the SSA to avoid retroactive application of new requirements. Any changes that impose new obligations on MA plans should be developed after appropriate and substantive interaction with the industry and apply only to payment years arising after the RADV proposal is finalized (i.e., prospectively). In other words, CMS can only apply changes in RADV methodology to payment years after publication of a final rule, and plans must have the ability to factor the RADV rules into their bids. Thus, even if CMS were to finalize any changes to the RADV methodology in 2019, the earliest it could apply would be 2021.

- **Acknowledge that a FFS adjuster is required under statute and improve CMS' audit methodology.** For many years, CMS expressly stated that a FFS adjuster is needed to meet statutory requirements for actuarial equivalence. CMS has reversed that long-held position with this rule. We strongly urge CMS to keep its word and develop a FFS adjuster, taking into account the multiple recent independent analyses finding such an adjustment is necessary, material, and legally required. In addition, the agency should improve the RADV audit methodology, including the design of a better process for determining whether a patient in fact has a given health condition through use of pertinent data sources.

- **Engage in meaningful, collaborative dialogue with plan partners to implement changes.** We urge CMS to create a fair and open process to develop appropriate payment oversight standards, similar to processes used in the traditional Medicare program and in certain aspects of the Food and Drug Administration's oversight of the pharmaceutical industry. This is especially important given the complexities of the MA payment system where various components (from benchmark-setting to risk adjustment to oversight) determine payments. Adopting such an approach is critical to the continued strength of the MA program and the ability of plans to meet the needs of the people they serve. The industry stands ready to work closely and collaboratively with CMS on the issues described here and other matters related to oversight of the MA program.

Conclusion

The RADV proposal violates numerous statutory requirements and is fundamentally unfair and ill-conceived. We urge CMS in the strongest possible terms to withdraw it and establish a collaborative process with stakeholders to create a workable alternative. We look forward to providing any additional information you may need and to continuing to work together to improve the health of the millions of Medicare beneficiaries our members serve.

Sincerely,

Matthew Eyles

Matthew Eyles
President and CEO

Enclosure

**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 ("Proposed Rule")**

I.   **Summary: The Centers for Medicare & Medicaid Services' (CMS) proposed changes to
     the Risk Adjustment Data Validation (RADV) program fail to satisfy the Social
     Security Act, are based on flawed data, and are procedurally defective**

The Proposed Rule includes several critical changes relating to RADV audits:

- CMS would extrapolate RADV audit findings to Medicare Advantage (MA) contract level
  payments.
- The extrapolation would be applied retroactively, going back to audits from 2011 and
  after.
- RADV audits for payment years 2011, 2012, and 2013 would be extrapolated based on a
  methodology described in a notice dated February 24, 2012 posted on the CMS website[1]
  (the 2012 RADV Notice).
- Audits for payment years 2014 and beyond would also be extrapolated. CMS has
  subsequently noted that it will extrapolate some, but not all, of the 2014 audits. In the
  Proposed Rule, CMS indicates it could use a different methodology for extrapolation
  including a potential approach that would use sub-cohorts of enrollees and has
  subsequently indicated that a sub-cohort methodology will be used for the 2014 audits.
- In the 2012 RADV Notice, CMS stated that it would apply a fee-for-service (FFS) adjuster
  as an offset to any RADV recovery amount. CMS also stated that it would conduct a study
  to determine the amount of the FFS adjuster. Specifically, in the 2012 RADV Notice,
  CMS stated: "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)." However, under the Proposed Rule, CMS reverses course by stating that a FFS
  adjuster is not appropriate as an offset for RADV recoveries.
- CMS states that based on a technical analysis of "audit miscalibration error" in the risk
  adjustment model (the CMS technical study[2]), a FFS adjuster is unnecessary because the
  impact of audit miscalibration is negative and extremely close to zero. Separately, CMS
  asserts that a FFS adjuster is not appropriate, regardless of what the study found, because

---

[1] Notice of Final Payment Error Calculation Methodology for Part C Medicare Advantage Risk Adjustment Data
Validation Contract-Level Audits, accessed at: 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.
[2] Fee for Service Adjuster and Payment Recovery for Contract Level Risk Adjustment Data Validation Audits -
Technical Appendix, available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-
Programs/Medicare-Risk-Adjustment-Data-Validation-Program/Other-Content-Types/RADV-Docs/FFS-Adjuster-
Technical-Appendix.pdf.

it would only correct payments to audited plans, which would be inequitable to unaudited plans.

- Following the publication of the Proposed Rule, CMS published additional data regarding the technical study and also released an addendum that contained additional information on the study's assumptions and methodology.[3]

Below we highlight our strong policy and legal objections to each component of the RADV proposal, beginning with the FFS adjuster. We have attached an analysis from the actuarial firm Milliman[4], based on CMS' data and methodologies as presented in the CMS technical study and the addendum to the study, that shows a FFS adjuster, or other similar adjustment, is necessary to ensure actuarial equivalence between payments to MA plans and payments under the FFS program, as required by the Social Security Act (SSA). Milliman finds that the CMS technical analysis "cannot appropriately be used to conclude a FFS adjuster is not required." Milliman adjusts for errors in CMS' methodology to find that an FFS adjuster is both positive and material. As such, Milliman's study refutes CMS' conclusion that a FFS adjuster is not necessary. Another study highlighted below, and two recent federal district court decisions, also conclude that a FFS adjuster is required.

In addition, CMS' separate argument that it would be inequitable to have a FFS adjuster is a distraction from the question at hand – which is whether or not a FFS adjuster is required to ensure actuarial equivalence. The issue of actuarial equivalence arises whenever CMS seeks to apply a documentation standard for payment (medical records) that differs from the standard used in developing the risk adjustment model (claims). Plans not subject to a RADV audit may still be subjected to different document standards if they face overpayment claims by the government, and two district courts have recently held that the actuarial equivalence requirement must be satisfied in the overpayment context. Thus, equity in fact requires the application of a FFS adjuster to RADV audits. Further, unaudited plans that do not face payment recovery issues are not adversely affected by different documentation standards, and therefore the use of a FFS adjuster would not adversely affect them. In any event, as a membership organization representing the MA industry, we can say, without hesitation, that our members support a FFS adjuster regardless of whether one of their contracts is selected for audit.

The proposal also includes several other substantive and procedural defects. For example, CMS does not have statutory authority to use its proposed extrapolation methodology. Even if it did, CMS' published 2012 extrapolation methodology is so flawed that implementation would be arbitrary and capricious. (We have also attached an analysis from the actuarial firm Wakely that identifies the significant concerns with the 2012 RADV methodology.)

Further, CMS' proposed retroactive application of the regulation going back almost a decade – to audit results for the 2011 payment year, based on diagnosis data from 2010 – is impermissible under the law. In addition, CMS proposes to apply its 2012 published extrapolation methodology

---

[3] Addendum to the Fee-for-Service Adjuster study, available at: 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.
[4] Available at: http://us.milliman.com/insight/2019/Medicare-Advantage-RADV-FFS-adjuster/.

to 2011-2013 audits using an extrapolation methodology never developed through rulemaking. CMS is also moving forward with a new methodology for audits of the 2014 plan year without providing adequate detail or any opportunity for comment, despite statutory rulemaking requirements. These actions are inconsistent with the SSA requirement for notice-and-comment rulemaking as indicated by the U.S. Supreme Court in its recent ruling in *Azar v. Allina Health Services*.[5]

Given the substantive and procedural defects in CMS' proposals, we urge CMS to take the following steps:

- Withdraw these proposals in their entirety.
- Affirm that the agency cannot change regulations retroactively.
- Acknowledge that a FFS adjuster is required under statute to ensure actuarial equivalence since documentation standards applied to payment differ from the standards used in developing the risk adjustment model.
- Improve the agency's audit methodology.
- Note that CMS does not have statutory authority to conduct extrapolations under RADV.
- Engage in meaningful, collaborative dialogue with the industry to address these issues going forward.

II. FFS adjuster is required by law

The Medicare statute clearly requires actuarial equivalence in payments between traditional Medicare and the MA program.[6] Actuarial equivalence can either be achieved through a FFS adjuster in assessing payment errors in the MA program, or, alternatively, through CMS estimating the risk adjustment model using audited FFS data. This interpretation was supported by CMS itself and has been upheld by two recent court decisions. The Milliman study shows that a FFS adjuster, or other similar adjustment, is necessary to ensure actuarial equivalence between payments to MA plans and payments under the FFS program. In addition, CMS' claim that equity concerns would prevent application of a FFS adjuster is meritless. Each of these issues are discussed in detail below.

A. FFS adjuster is required by statutory language ensuring actuarial equivalence

i. Background on statute and risk adjustment

Section 1853(a)(1)(C)(i) of the Social Security Act (SSA) states in relevant part that the Secretary of HHS:

> shall adjust the payment amount [to an MA plan] for such risk factors as age, disability status, gender, institutional status, and such other factors as the Secretary determines to be appropriate, including adjustment for health status . . . , so as to ensure actuarial equivalence. The Secretary may add to, modify, or substitute for

---

[5] 139 S.Ct. 1804 (2019).
[6] See Section 1853(a)(1)(C)(i) of the SSA.

such adjustment factors if such changes will improve the determination of actuarial equivalence.

CMS makes the required statutory adjustments for health status through a risk adjustment model that applies a risk score to the MA payment for each enrollee. A risk score represents the relative costs of an individual compared to that for an average beneficiary. CMS uses the CMS-Hierarchical Condition Category (HCC) model to calculate these risk scores, which are determined based on demographic (e.g., age, gender, Medicaid status) and disease characteristics. Diseases are assigned to HCCs. In this sense, each HCC represents a disease group (e.g., diabetes, congestive heart failure). In general, each plan is paid its bid, multiplied by the risk score for that enrollee. The bid represents the average costs for an enrollee in that plan to receive Medicare Parts A and B items and services, standardized to a 1.0 risk score. The risk model pays more for sicker enrollees, and less for healthier enrollees.

The CMS-HCC risk model is estimated – or calibrated – on claims data from traditional Medicare (also known as Medicare FFS). Diagnoses from a previous year are used to estimate costs in a current year (e.g., 2013 diagnoses to predict 2014 costs). CMS uses a weighted least squares regression to determine the dollar amount associated with each HCC (e.g., chronic obstructive pulmonary disease, diabetes, etc.) and for demographic characteristics (e.g., age, gender, etc.). These estimated dollar amounts for each HCC or demographic characteristic are also known as model coefficients.

Through a process that CMS describes as "normalization", CMS runs the model on FFS claims data to determine the total predicted spending for the population. CMS divides the total predicted spending by the number of enrollees to determine the average predicted spending for the population. CMS then converts the dollar amounts for the model coefficients to relative factors by dividing each coefficient by the average predicted costs for the population. The sum of these relative factors is the beneficiary risk score. This process leads to an average risk score of 1.0, because all dollar coefficients are divided by the average predicted spending for the population.

As noted earlier, plans submit bids to CMS that estimate the expected costs for their population to provide Medicare Part A and B items and services. For example, if the plan's standardized bid is $1,000 per member per month, and the risk score for an enrollee is 1.1, the plan will be paid $1,100 for that enrollee.

CMS conducted audits for plan years 2011, 2012, and 2013 to determine whether diagnoses submitted by plans and used by CMS to determine risk scores were supported by medical records.[7] CMS used a documentation standard for these audits that differed from the documentation standard used to develop the risk adjustment model. The standard CMS used to determine payment errors under a RADV audit of an MA plan was the medical record, while the documentation standard used to develop the relative values in the HCC model was unaudited FFS claims data. This different documentation standard is why a FFS adjuster is necessary.

---

[7] CMS also conducted pilot and contract-level audits for plan year 2007.

ii.   Actuarial experts and court decisions affirm the need for a FFS adjuster to ensure
        actuarial equivalence

The Actuarial Standards Board's Actuarial Standards of Practice (ASOP) provide guidelines for an actuarial review of risk adjustment models. These standards expressly provide that "[t]he type of input data that is used in the application of risk adjustment should be reasonably consistent with the type of data used to develop the model."[8] ASOP No. 45 requires consistency between how the model is developed and how it is applied in payment. However, the documentation standards used in the RADV audit to determine a plan's payment error are different from the documentation standards used to develop the risk adjustment model.

In addition, the United States District Court for the District of Columbia recently examined the underlying principle of actuarial equivalence and its applicability to section 1853(a)(1)(C)(i)'s actuarial equivalence requirement in a clearly written and well-reasoned opinion. *See UnitedHealthcare Ins. Co. v. Azar*, 330 F. Supp. 3d 173 (D.D.C. 2018) (Collyer, J.), *appeal docketed*, No. 18-5326 (D.C. Cir. Nov. 14, 2018).[9]

The *UnitedHealthcare* court found that section 1853(a)(1)(C)(i) imposes a non-discretionary duty on CMS whereby the agency must ensure that "two modes of payment" – payment to providers and suppliers under Medicare Parts A and B, on the one hand, and payment to MA plans under Medicare Part C, on the other – result in "present values [that] are equal under a given set of actuarial assumptions." 330 F. Supp. 3d at 185 (citations omitted). And by "given," the *UnitedHealthcare* court meant "'the same,' as in two figures are actuarially equivalent when they share the same set of actuarial assumptions." Id. at 185-86. And that "[d]ifferent assumptions behind the elements of a calculation would, *necessarily*, result in actuarially non-equivalent results." Id. at 186 (emphasis added). In other words, the assumptions used in Medicare FFS payments *must* hold true for payments to MA plans.

Applying the plain meaning of what constitutes "actuarial equivalence", the *UnitedHealthcare* court vacated the CMS final rule on overpayments promulgated in 2014. A critical part of the holding was that CMS violated the actuarial equivalence requirement by calculating risk adjustment payments to MA plans using unsubstantiated FFS diagnosis codes to determine the expected additional costs of providing coverage to a beneficiary with a particular medical condition, while holding MA plans to a standard of perfection whereby diagnosis codes on claims submitted by MA plans had to be absolutely free of diagnosis error. The court found that "CMS cannot subject the diagnosis codes underlying [MA] payments to a different level of scrutiny than it applies to its own payments under traditional Medicare without impermissibly skewing the

---

[8] Actuarial Standard of Practice No. 45 § 3.2.

[9] However, despite the significance of this decision, the Proposed Rule makes only passing mention of it in a footnote that states: "We are aware of the district court's recent ruling in *United HealthCare Insurance Co.* v. *Azar*, No. 16-cv-157 (D.D.C. September 7, 2018), and the government is reviewing that decision and considering its response. In any event, that ruling was made on the basis of the administrative record before the court, which did not include the results of our study." 83 Fed. Reg. at 55,040 n.29. A few days after the Proposed Rule was published, the government filed a motion for reconsideration in the district court citing the study and claiming that it constitutes "new evidence." *See* Defs.' R. 60(b) Mot. for Partial Recons., *UnitedHealthcare Ins. Co. v. Azar*, No. 1:16-cv-00157 (D.D.C. Nov. 5, 2018), ECF No. 76.

calculus: by doing so, it ensures that there will *not* be actuarial equivalence between traditional Medicare payments and [MA] payments for comparable patients." Id. at 186. The court took particular note of the fact CMS had acknowledged this very principle in 2012 when it promised that in determining the final payment recovery amount in RADV audits, CMS would "apply a Fee-for-Service Adjuster (FFS Adjuster) amount as an offset to the preliminary recovery amount." Id. at 188.

That is, if the documentation standard used in determining payments to MA plans is not the same as the standard used to calibrate risk model coefficients based on FFS claims data, there must be an adjustment. Alternatively, CMS could achieve actuarial equivalence by estimating the risk adjustment model using audited FFS data. However, by eliminating the FFS adjuster and continuing to estimate the risk model based on unaudited FFS claims data, CMS would hold MA organizations to a perfection standard for medical record documentation that is clearly not applied in the FFS program – as shown in the error rates by HCC that CMS identified in its own study. To do so would contravene the SSA's "actuarial equivalence requirement."

In another recent decision, the United States District Court for the Central District of California also examined the statutory provision in section 1853(a)(1)(C)(i) on actuarial equivalence. *See United States ex rel. Benjamin Poehling v. UnitedHealth Group, Inc.*, No. 16-08697, 2019 WL 2353125 (C.D. Cal. Mar. 28, 2019). The government argued that the statutory language "merely arms the Secretary with broad *discretionary* power to adjust payment levels based on the health status of Medicare beneficiaries." Id. at 5 (emphasis in original). In denying the government's motion for partial summary judgment, the court stated that it was "unpersuaded by the Government's argument in light of the plain language of the statute, which provides that the Secretary *shall* adjust the payment amount for factors the Secretary deems appropriate *so as to ensure* actuarial equivalence. Such language is far from discretionary." Id. at 6 (emphasis in original). The court also cited the decision in *UnitedHealthcare Ins. Co. v. Azar* on the need for a FFS adjuster as "persuasive authority". Id.

iii. Independent actuarial analyses clearly demonstrate the need for a FFS adjuster

Since CMS published the Proposed Rule, two independent analyses – one by Avalere (the Avalere study[10]) and one by Milliman (the Milliman study) – clearly demonstrate the need for a FFS adjuster. Importantly, neither analysis estimates the appropriate amount of the FFS adjuster. Rather, each study demonstrates that CMS' methodology, after adjusting for methodological flaws, would lead to a FFS adjuster that is *not zero*. Thus, CMS' conclusion – that the audit miscalibration error is negative and extremely close to zero and therefore a FFS adjuster is not necessary – cannot be supported by its own technical study.[11]

---

[10] Avalere Health. Eliminating the FFS Adjuster from the RADV methodology may affect plan payment. March 2019. Available at: https://avalere.com/wp-content/uploads/2019/03/20190318-FFS-Adjuster-Analysis-Final-.pdf.

[11] The methodological flaws demonstrate that, at a minimum, CMS has "relied on factors which Congress has not intended it to consider, entirely failed to consider an important aspect of the problem, offered an explanation for its decision that runs counter to the evidence before the agency, or [proposed a course of action that] is so implausible that it could not be ascribed to a difference in view or the product of agency expertise." *Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 43 (1983).

a. Avalere study

Avalere analyzed CMS' conclusion in the Proposed Rule that re-estimating risk scores from the HCC model, based on coding errors in FFS claims, does not have an impact on risk scores. According to CMS, the risk scores from the re-estimated model were almost equal to those produced by the original model. This led CMS to determine that erroneous coding in FFS has minimal impact on MA risk scores, and therefore no FFS adjuster is needed.

Avalere noted that "certain key assumptions embedded in CMS' analysis do not appropriately capture the full variation in the data and minimized the impact of documentation error." For example, Avalere explained that CMS' simplifying assumption that each person in the sample has an average number of claims supporting a particular disease group, or HCC, is flawed because the distribution of Medicare claims is skewed. That is, the average number of claims is higher than the median, or midpoint. Avalere found, just by applying CMS' methodology to the *actual* distribution of error rates, that MA risk scores from the re-estimated model accounting for coding errors in FFS claims would be almost 8 percent lower than the original model. Avalere also says that "assuming that each claim supporting an HCC has an equal probability of error suggests that coding and documentation errors occur randomly. However, it is probable that there are correlations in errors."

b. Milliman study

AHIP sponsored an analysis by Milliman to evaluate CMS' conclusion that a FFS adjuster is not necessary. Milliman reproduced CMS' methodology and used CMS' published assumptions, the data, and the related files CMS provided.[12]

The Milliman study identifies multiple significant issues in the CMS assumptions and methodologies. The Milliman study focused on and adjusted for significant shortcomings in CMS' assumptions relating to how: 1) CMS normalized the risk adjustment model, and 2) CMS derived and applied beneficiary-level diagnosis error rates. Milliman found that CMS did not state that it calculated a FFS adjuster in the technical analysis accompanying the Proposed Rule and that "the CMS analysis measured a model calibration difference rather than addressing the question of whether a FFS adjuster is required in RADV audits."

The study explains it did not attempt to identify all potential issues and makes no judgment about the appropriateness of other methodologies that could be used to determine an appropriate FFS

---

[12] These data, released by CMS in March and June 2019, include: diagnosis data used to calibrate the CMS-HCC model through 2011; the model calibration file used to calculate 2009 MA payments; medical record review findings from the 2008 Comprehensive Error Rate Testing review; a mapping of ICD-9 diagnosis codes to version 12 of the CMS-HCC model; MA data for a sample of RADV eligible and non-RADV eligible beneficiaries from the CMS Enrollment Data Base, Model Output File, and Monthly Membership Report for 2011; dollar coefficients and risk factors for the original data, as well as 50 simulated 'corrected' iterations of the data, both before and after an adjustment to account for deletion bias is made to each iteration; text file versions of the SAS programs used to conduct the analysis summarized in the study and addendum; and a variable crosswalk and sort file used in the program to conduct the analysis. Upon review of the published SAS code, Milliman verified that the CMS implementation of the process described in the technical appendix was not materially different from its reproduction of the CMS analysis.

adjuster. Milliman further notes that, 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 the paper. Milliman states in the study that "we have not been able to conceive of a reasonable methodology that would lead to the conclusion a FFS adjuster is unnecessary."

Milliman summarizes its findings in the Executive Summary of the study as follows:

> The Centers for Medicare and Medicaid Services (CMS) issued a proposed rule[13] 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[14]. 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." [15] 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:
>
>   - o CMS assumes diagnosis coding errors are independent from each other, which materially understates HCC error rates in FFS.
>
>   - o 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.

---

[13] 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).

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

[15] 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).

o 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 that a FFS adjuster is required.

Milliman further explains the purpose of their study as follows:

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 likely 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.

While the Milliman study does not determine the appropriate amount of a FFS adjuster, it includes an estimate of what the FFS adjuster would be if it were to be calculated using 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).[16] Milliman explains that:

Under this approach, we calculated a FFS adjuster using claim level error rates, actual distributions of the number of diagnoses (assuming full *independence[17]*), and an HCC error rate of 33% (assuming full *dependence*[18]), in addition to several scenarios in between. This approach resulted in estimated values of a FFS adjuster[19]

---

[16] Milliman's study is based on CMS' data and methodology. We discuss additional flaws with the CMS data and methodology that Milliman did not correct for in Sections II.B.ii-iv below.

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

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

[19] 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.

between 8% and 21%. For perspective, 8% of federal payments to MAOs exceeds $16 billion and 21% exceeds $42 billion per year,[20] 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.

iv.   Simplified illustrations of why actuarial equivalence requires a FFS adjuster

To see how actuarial equivalence works in practice, and why CMS violates actuarial equivalence if it does not apply a FFS adjuster, consider the following example. It is described in Table 1 and discussed in the Milliman study. This simplified example is based on an example that CMS developed when considering the need for a FFS adjuster.[21]

Assume CMS develops a risk model based on four individuals in FFS Medicare. In the example, the only cost of treatment is associated with diabetes. Each of the four is coded as having diabetes. The cost of treating a person with diabetes (which is supported in the medical record) is $4,000. The cost of a person who actually does not have diabetes (the medical record has no support for diabetes) is $0. Because CMS estimates the HCC model on diagnoses codes from

---

[20] 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.

[21] See Decl. of Daniel Meron, Ex. B at 8, *UnitedHealthcare Ins. Co. v. Azar*, No. 1:16-cv-00157 (D.D.C. Oct. 2, 2017), ECF No. 44-3.

claims, regardless of whether they are supported by the medical record, CMS divides the $12,000 of total cost by the count of beneficiaries with a diabetes diagnosis on a claim. In this example, because four beneficiaries have diabetes on the claims, the estimated payment to treat diabetes is $12,000 divided by 4, or $3,000. Importantly, beneficiary D in Table 1 below does not have diabetes coded on the medical record, but since it is included on the claim, that person is used to determine the payment for an individual with diabetes.

Table 1. Example Showing Calculation of MA Payment Amount for Diabetes

	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

Now consider the example illustrated in Table 2 below, which is also described in the Milliman study and based on an example from CMS. In this example, a plan has five enrollees who had diabetes coded on claims, but three of them have diabetes supported in the medical record, and two do not. The total cost for the five beneficiaries in FFS is $12,000. However, if CMS were to recover funds for unsupported codes in a RADV audit without a FFS adjuster, CMS would take back $6,000 (for Beneficiaries D and E). This means the plan would be paid only $9,000, which is $3,000 less than under FFS.[22] The example clearly demonstrates there is not actuarial equivalence between FFS and MA when a RADV audit is performed without a FFS adjuster.

Table 2. Example Showing Actuarial Equivalence Not Achieved

	Diabetes on Claim?	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

---

[22] As Milliman notes, 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 risk score of 1.0.

     v.   <u>Applied example demonstrating the need for a FFS adjuster that includes risk adjustment model calibration</u>

The Milliman study builds on the above example to show the need for a FFS adjuster within the MA payment framework. Below and in Table 3 are the key elements of the analysis.

For this example, Milliman created a simplified risk model using a least squares regression[23] that includes demographic and disease components, which is what CMS does when it estimates the CMS-HCC model. In this example, there are four individuals – two are 70 years old, one is 75, and one is 80 – and all have diabetes coded on their claims. Milliman estimated the model using the same set of assumptions that CMS uses, where only the claim is used as documentation.

Table 3. Model Estimated Based on Claim Information

	On Claim?	FFS Cost (Actual)	Predicted FFS Cost	Relative Coefficient
Beneficiary 1				
70 year old			$6,500	0.650
Diabetes	Yes		$3,000	0.300
Subtotal		$9,000	$9,500	0.950
Beneficiary 2				
70 year old			$6,500	0.650
Diabetes	Yes		$3,000	0.300
Subtotal		$10,000	$9,500	0.950
Beneficiary 3				
75 year old			$7,000	0.700
Diabetes	Yes		$3,000	0.300
Subtotal		$10,000	$10,000	1.000
Beneficiary 4				
80 year old			$8,000	0.800
Diabetes	Yes		$3,000	0.300
Subtotal		$11,000	$11,000	1.100
Total		$40,000	$40,000	1.000

Milliman then applies these figures to a case in which a plan has four individuals with a claim of diabetes, but only three have the diagnosis supported in the medical record. This example assumes a plan bid of $10,000 per year. See Table 4. As discussed further in the attached Milliman report, without a FFS adjuster, actuarial equivalence will not be achieved because plan payments will be $37,000 when the actuarially equivalent amount is $40,000.

---

[23] As noted in the Milliman study, due to the simplistic nature of this example, the least squares regression does not produce a unique solution. Milliman 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.

Table 4. RADV With and Without a FFS Adjuster

	On Claim?	On Medical Record?	Coefficient	MA Payment Without FFS Adjuster		MA Payment With FFS Adjuster	
				Before RADV	After RADV	Before RADV	After RADV
Beneficiary 1							
70 year old			0.650	$6,500	$6,500	$6,500	$6,500
Diabetes	Yes	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 year old			0.650	$6,500	$6,500	$6,500	$6,500
Diabetes	Yes	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 year old			0.700	$7,000	$7,000	$7,000	$7,000
Diabetes	Yes	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 year old			0.800	$8,000	$8,000	$8,000	$8,000
Diabetes	Yes	No	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

The Milliman study includes several additional scenarios that review the impact of calibrating the risk model with one set of documentation standards yet recovering funds in RADV audits using a different set of documentation standards. All these examples demonstrate that when the CMS-HCC model is calibrated and normalized based on unaudited claims data, a FFS adjuster is necessary under a RADV audit to maintain actuarial equivalence as required by statute.

      vi.   <u>Nothing in SSA sections 1853(a)(1)(C)(ii) or (iii) change the requirement for a FFS adjuster.</u>

In the request for additional comment published in the Federal Register on June 28, 2019, CMS sought input "on whether section 42 U.S.C. 1395w–23 [section 1853 of the SSA]—and in particular clause (a)(1)(C), which requires risk adjustment in subclause (a)(1)(C)(i), mandates a downward adjustment of risk scores in subclause (a)(1)(C)(ii), and includes provisions about risk adjustment for special needs individuals with chronic health conditions in subclause (a)(1)(C) (iii)—mandates an FFS Adjuster, prohibits an FFS Adjuster, or should otherwise be read to inform our proposal not to apply an FFS Adjuster in any RADV extrapolated audit methodology."

As stated above, section 1853(a)(1)(C)(i) clearly requires actuarial equivalence in payments between traditional Medicare and the MA program.[24] Actuarial equivalence can either be achieved through a FFS adjuster in assessing payment errors in the MA program, or, alternatively, through CMS estimating the risk adjustment model using audited FFS data. This interpretation was supported by CMS itself and has been upheld by two recent court decisions. The provisions in subsections (ii) and (iii) relate to adjustments for coding intensity and for risk adjustment for new enrollees in chronic condition special needs plans. They have nothing to do with the requirement for a FFS adjuster under subsection (i). For example:

- Under the plain language of the statute, subsection (i) does not refer to subsections (ii) or (iii). The requirements are completely independent. Congress added subsections (ii) and (iii) a number of years after subsection (i), but despite multiple chances to change the actuarial equivalence language in subsection (i), did not do so. Thus, there is nothing in the statute to suggest subsections (ii) and (iii) support removal of the FFS adjuster from RADV methodology.
- The coding intensity adjustment in subsection (ii) addresses coding pattern differences between MA and FFS. CMS has expressly stated that RADV audits address coding *accuracy* issues, not coding pattern differences.[25] Accordingly, even if the statute gave CMS discretion to not apply a FFS adjuster based on provisions in subsection (ii) (which it does not), CMS could not avoid applying a FFS adjuster without explaining its shift in legal interpretation that the two provisions are unrelated; providing a detailed analysis demonstrating how the coding intensity adjustment allegedly undercut the need for a FFS adjuster; and providing a comment opportunity through formal notice-and-comment rulemaking.
- Subsection (iii) requires CMS to apply a higher risk score for new enrollees in special needs plans for those with chronic conditions. This provision is clearly irrelevant to the general actuarial equivalence requirement in subsection (i) or the need for a FFS adjuster.

---

[24] See Section 1853(a)(1)(C)(i) of the Social Security Act.
[25] Announcement of Calendar Year (CY) 2019 Medicare Advantage Capitation Rates and Medicare Advantage and Part D Payment Policies, at 38-39 (April 2, 2018).

B. CMS' study fails to support its position on the FFS adjuster

Notwithstanding the *UnitedHealthcare* decision and its reliance on the 2012 RADV Notice, CMS now proposes to eliminate the FFS adjuster and offers two flawed and unsupportable reasons for doing so:

- Systematic Effect: Ignoring the agency's previous conclusion that the FFS adjuster "accounts for the fact that the documentation standard used in RADV audits . . . is different from the documentation standard used to develop the Part C risk adjustment model," the Proposed Rule relies on the results of the CMS technical study. According to CMS, the study results "suggest[] that errors in FFS claims data do not have any systematic effect on the risk scores calculated by the CMS-HCC risk adjustment model, and therefore do not have any systematic effect on the payments made to MA organizations."
- Inequities Between Audited and Unaudited Plans: CMS also asserts that "even if [it] had found that diagnosis error in FFS claims data led to systematic payment error in the MA program, we no longer believe that a RADV-specific payment adjustment would be appropriate. . . . Doing so would introduce inequities between audited and unaudited plans, by only correcting the payments made to audited plans."

Below we discuss the limitations in the CMS technical study at length. In general, the CMS technical study fails to address the key issue of actuarial equivalence in the context of RADV audits. In addition, as the Milliman study points out, the level of the FFS adjuster depends in large part on the assumptions used. However, the CMS technical study contains multiple flaws and questionable assumptions that led to the calculation of artificially low error rates and, as a result, to CMS concluding that a FFS adjuster was not necessary. We reference the findings from the Milliman study where appropriate in each section on these limitations. After our discussion on the limitations of the CMS technical study, we discuss our strong disagreement with the rationale used to justify the inequities argument.

The limitations in CMS' study are as follows:

  i. Limitation #1: An analysis of the systematic effects of the risk model does not address the actuarial equivalence question in the context of RADV audits

The question CMS identified in the 2012 RADV Notice related to how the FFS adjuster applies within the context of RADV, when recoveries are made if a medical record does not support a diagnosis code, but the risk model is developed based on claims and medical records that are not reviewed. This is the same issue addressed in the *UnitedHealthcare* case, where the court noted that "two figures are actuarially equivalent when they share the same set of actuarial assumptions. Different assumptions behind the elements of a calculation would, necessarily, result in actuarially non-equivalent results." *UnitedHealthcare*, 330 F. Supp. 3d at 186.

The Proposed Rule, however, makes no effort to address this meaning of "actuarial equivalence" in the context of RADV. As discussed in Section II.A.vi above, CMS did not even seek public comment on whether the statutory language mandating actuarial equivalence at section 42 U.S.C.

1395w–23(a)(1)(C) should be considered in the context of applying a FFS adjuster in the RADV audit methodology until the agency released additional data in June 2019. Instead, the Proposed Rule, through the CMS technical study, purports to ask and answer a different question: namely, whether diagnosis errors in FFS claims have a "systematic effect on the risk scores calculated by the CMS-HCC risk adjustment model, and therefore [a] systematic effect on the payments made to MA organizations."

This issue is irrelevant to the question of whether a FFS adjuster is needed to ensure actuarial equivalence in the context of RADV audits. That is, even assuming for the sake of argument that the CMS study actually supports the aggregate negative "systematic effect" conclusion for which it is cited (which in fact it does not), section 1853(a)(1)(C)(i) of the SSA imposes a mandatory duty on CMS to ensure actuarial equivalence. This mandatory duty was recognized in recent rulings by district courts in both the *UnitedHealthcare Ins. Co. v. Azar* and *Poehling* proceedings, as described above.

This requirement does not terminate when payment is made to an MA plan. The statute's use of the word "ensure" confirms that the actuarial equivalence requirement remains in effect through and including whatever post-payment audit process CMS may devise. The actuarial equivalence requirement is not met when CMS estimates the model on unaudited FFS data because it uses an audit methodology that applies a documentation standard drastically different from that applied to FFS claims. The use of these different documentation standards in model estimation and payment must therefore be addressed through the application of a FFS adjuster. Not doing so violates the fundamental standards of actuarial equivalence, which require consistency between the way a risk adjustment model is developed and how it is applied.

The continuing nature of the statutory duty imposed by section 1853(a)(1)(C)(i) of the SSA was at the heart of the decision in *UnitedHealthcare*. CMS stated before the court that the statute only imposed a duty regarding the manner in which the agency calculated initial payments made to MA plans. However, as Judge Collyer explained when questioning counsel for CMS during oral argument: "Their argument [referring to the *UnitedHealthcare* plaintiffs] is that by figuring the coefficients on unaudited files and then paying out but requiring repayment on anything that is not substantiated in a medical record is to start at the beginning as if it were actuarially equivalent, but set up a system whereby it no longer is. The beginning is arguably equivalent, the process is not." Hr'g Tr. 39:14–24, *UnitedHealthcare Ins. Co. v. Azar*, No. 1:16-cv-00157 (D.D.C. Aug. 9, 2018), ECF No. 73.

Therefore, even if errors in diagnosis coding under Medicare Parts A and B do not have a "systematic effect" on the aggregate risk scores used to calculate payments made to all MA plans, section 1853(a)(1)(C)(i) of the SSA requires CMS to take into account those diagnosis coding errors when determining how much, if anything, a particular MA plan may owe as the result of a RADV audit. This requires considering the particular enrollees included in the sample under review and, if contract-level extrapolation is deemed lawful, the particular enrollees included in the MA contract under review. The statute permits no construction—let alone a reasonable construction—that applies a different standard. *See, e.g., Michigan v. EPA*, 135 S. Ct. 2699, 2708 (2015) ("*Chevron* allows agencies to choose among competing reasonable interpretations of a

statute; it does not license interpretive gerrymanders under which an agency keeps parts of statutory context it likes while throwing away parts it does not.").

  ii. Limitation #2: The CMS technical study is based on inconsistent and arbitrary data

CMS uses three sources of input data for its technical study:

1) Calendar Year (CY) 2008 Comprehensive Error Rate Testing (CERT) audit data
2) 2004-2005 FFS claims data
3) Two million MA records sampled from the 2011 overpayment run, split evenly between RADV eligible and non-RADV eligible beneficiaries[26]

CMS does not explain why it chose each of these sources, in terms of the data or the time periods which the data represent. However, the data and time periods raise numerous questions. For example:

- More recent FFS claims data and MA records could have been used, rather than 2004-2005 FFS claims and 2011 MA data.
- By conducting its analysis using data sources from different years, the study may be inappropriately accounting for differences in health care treatment patterns between these different time periods.
- One-half of the MA records relate to beneficiaries who *are not eligible* to be included in RADV audits, which raises serious questions about whether the study could accurately assess the impact of removing diagnosis errors in a RADV audit of beneficiaries who *are eligible* to be included.
- CMS acknowledges on page 55037 of the Proposed Rule that the CMS-HCC model is "recalibrated approximately every 2 years to reflect newer treatment and coding patterns in Medicare FFS." Given this practice, it is unclear why CMS is relying on a point-in-time analysis of coding patterns from over 10 years ago.

The absence of any explanation for these data sources raises transparency concerns as commenters are left without an indication of the rationale for CMS' decision to use these data.

  iii. Limitation #3: The CMS technical study is not a RADV-like review

In the 2012 RADV Notice, CMS states that the amount of the FFS adjuster would be "calculated by CMS based on a RADV-like review of records submitted to support FFS claims data." However, CMS did not conduct a RADV-like review of FFS claims data. In a RADV audit, CMS

---

[26] Per Section 1128J(d) of the SSA and the overpayment regulation 42 CFR §422.326, all MA plans are required to report and return overpayments. CMS recovers these overpayments on an annual basis by conducting "risk score reruns" for prior payment years within a six-year lookback period. From the data subsequently released by CMS in March 2019, we understand that CMS sampled MA records from those submitted for PY 2011 (2010 dates of service). However, CMS does not clarify that these data had been processed for overpayment recovery or specify during which calendar year the 2011 overpayment run took place (the most recent deadline for PY 2011 overpayment submission was July 6, 2018 and risk scores for 2011 were rerun for overpayment recovery purposes in payments made to MA plans on October 1, 2018).

randomly selects 201 beneficiaries from a MA contract in equal numbers across three strata based on risk score, determines whether the HCCs submitted for payment for each of these beneficiaries is substantiated based on a prescriptive medical record review, and then determines a payment error based on the difference between the original risk scores calculated for these beneficiaries and the corrected risk scores based on the medical record review. This payment error would then be extrapolated to determine a final recovery amount.

Instead of following the RADV methodology in reviewing FFS claims data, CMS designed its flawed study to determine the impact on the CMS-HCC risk model of diagnosis coding errors in FFS claims data. In addition, the RADV parameters were not strictly followed. For example:

- Instead of measuring diagnosis discrepancies using data sampled at the beneficiary level, CMS generated an estimate of these errors by using data sampled at the claims level. In its methodology, CMS commits a number of errors in assigning the beneficiary-level error rate based on this review of the sampled claims data. As a result, CMS' methodology is fatally flawed and does not represent an accurate representation of the beneficiary-level error rates.
  - Specifically, instead of selecting a random sample of FFS beneficiaries and reviewing the medical records to support each risk adjustment-eligible diagnosis reported for those beneficiaries, CMS reviewed a random sample of 8,630 FFS outpatient claims from the CERT audit data.
  - The CERT data do not resemble a RADV audit sample in any way, and CMS also admits that these data lack a large enough sample size for many HCCs to generalize error rates to the total population.
  - If the CERT audit data were somehow shown to be appropriate, which has not occurred, 2008 data predates the implementation of ICD-10 codes and therefore have questionable applicability to diagnosis coding trends in today's environment.
- CMS includes beneficiaries who are not eligible for inclusion in a RADV study. Given that one-half of the MA records used in the CMS technical study were not RADV eligible, the study does not represent a 'RADV-like' review of claims data, as CMS stated it would conduct to generate the FFS adjuster.
- The CMS methodology does not take into account the process plans must follow to validate HCCs in RADV medical record review, which allows submission of a specified number of medical records to substantiate an HCC.
- As noted in its addendum to the technical study, CMS "excluded claims where providers refused to submit medical records, or did not provide sufficient documentation" rather than determining the diagnoses on these claims were not supported, as would have occurred in a RADV audit.
- Instead of comparing original *risk scores* from a sample of beneficiaries to corrected risk scores based on medical record review, CMS calculated *coefficients* based on FFS claims data reflecting the simulated error rates and applied those coefficients to MA data.

As the technical analysis that CMS performed in no way resembled a "RADV-like review" – which the agency stated it would conduct in order to calculate the amount of the FFS adjuster –

and relies on statistical concepts not relevant to the key issue of actuarial equivalence, the study cannot be used to support a position that the FFS adjuster is not necessary.

> iv.   Limitation #4: CMS makes invalid statistical assumptions about claim independence

A critical assumption that allows CMS to find that a FFS adjuster is not necessary is that each Medicare claim is independent of one another. By making this assumption, CMS dramatically understates the likelihood that a beneficiary will have a diagnosis coding error corresponding to a given HCC. For example, in its review of the 2008 CERT data, CMS finds that among HCC 80 (Congestive Heart Failure), 156 claims out of 519 claims were discrepant, for an error rate of 30.1 percent. The agency further notes that an average enrollee with HCC 80 would have six claims, which leads to a probability that HCC 80 would be in error for a beneficiary of $0.301*6.1=0.8$ percent. Not surprisingly, when CMS applies such low error rates to its data, the impact of removing diagnosis codes is minimal.

On page 9 of the CMS technical study appendix, CMS asserts – without any additional support – that "each enrollee HCC potentially has multiple claims with independent supportive medical records." In reality, coding errors will not be independent from one claim to the next, especially if the patient is seeing the same healthcare provider.

As pointed out in the Milliman study: "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 to at least some degree. The assumption that providers code randomly must hold to assume independence." Further, Milliman points out: "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." Avalere makes a similar point in its study, noting that "it is probable that there are correlations in errors. For example, a healthcare provider submitting multiple claims for the same beneficiary might repeat the same coding or documentation error."

To summarize, CMS cannot multiply probabilities when the events are not independent. If the same provider has seen the enrollee, it is far more likely that the events are dependent. If there is a 50 percent chance that a provider will make an error when seeing an enrollee, that same probability applies regardless of how many visits the enrollee has to the provider. This assumption by CMS – which Milliman critiques in their analysis – is simply not credible.

> v.   Limitation #5: The average number of claims per beneficiary cannot be used to determine a beneficiary level error rate

Milliman notes that using an average number of claims per member in calculating a beneficiary level error rate ignores the fact that the number of claims per person will vary. Avalere makes a similar argument in its study.[27] That is, some beneficiaries have more than the average number of

---

[27] Op cit. 9

claims, and some have fewer. In addition, the data are not normally distributed, as noted in the Avalere study. That is, a small number of enrollees have a large number of claims, which skews the distribution of claims underlying the CMS study.

Milliman describes an example using HCC 55 (Major Depressive, Bipolar, and Paranoid Disorders), which CMS assumes has about a 50 percent error rate on a HCC basis. Milliman presents an illustrative example where two beneficiaries have HCC 55; one has two claims, while the other has 10. In this case, the error rate, assuming independence of diagnosis coding errors, is $0.5 ^\wedge 2 = 0.25$ for the first beneficiary and $0.5 ^\wedge 10 = 0.0001$ for the second. By averaging these error rates, Milliman finds the average error rate is 0.125. By contrast, CMS' methodology, which focused on an average number of claims for all beneficiaries, would result in an average error rate of $0.5 ^\wedge 6 = 0.016$. In other words, Milliman finds an average error rate in this example that is nearly eight times higher than the error rate calculated using CMS' approach.

In Milliman's analysis, they adjust for this methodological error by using the actual distribution of claims, rather than the average number of claims. Due to limitations in the data provided by CMS, Milliman used the 5 percent Limited Data Set (LDS) claims files for this analysis. The Avalere study also used the LDS files. Milliman's and Avalere's use of the actual number of claims represents a much more accurate depiction of how error rates can be applied to determine a FFS adjuster. By using the average number of claims instead of the actual number of claims, CMS calculated an inaccurate estimate of the audit miscalibration. Making this adjustment, as Milliman demonstrates, dramatically increases the level of the necessary FFS adjuster and proves that CMS' conclusion that a FFS adjuster is not necessary is flawed.

      vi.   <u>Limitation #6: CMS does not properly "normalize" the risk adjustment model in its simulations</u>

In its methodology, CMS excludes unsupported diagnosis codes in the calibration of the CMS-HCC model. However, when transforming the coefficients calculated by the risk model to relative factors used to determine a risk score – a process referred to as normalization, which ensures the overall risk score is 1.0 – CMS includes the unsupported diagnosis codes and therefore does not correctly normalize the model. As Milliman finds, this step leads CMS to its erroneous conclusion that a FFS adjuster is not necessary, when in fact adjusting for this error alone demonstrates that a FFS adjuster is necessary. In addition, the magnitude of the FFS adjuster is material (that is, non-zero) – even when using the average number of claims, which as noted above is not a valid assumption.

CMS, in their Addendum to the FFS study[28], provided additional information on the normalization process. This additional detail, however, does not correct for any of the flaws inherent in the agency's chosen methodology. In particular, CMS includes a mathematical "explanation" of their approach to "offset deletion bias". CMS also described the Inflated Post-Audit Risk Score (IPARS) adjustment that they made to "offset the bias that the deletion procedure itself creates in expenditures."

---

[28] Op cit. 3

Milliman reviewed the mathematical arguments put forward by CMS in the Addendum and found notable flaws in CMS' approach. In particular, Milliman notes the following:

> 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.

With respect to the IPARS, Milliman finds that CMS' calculation of the IPARS is itself evidence of the need for a FFS adjuster. IPARS represents an interim step in the methodology that was implicit in the description CMS provided in its technical appendix but made explicit in the Addendum – IPARS is the name CMS ascribes to the process through which the CMS-HCC model was normalized inconsistently in order for CMS to conclude that a FFS adjuster is not necessary. In particular, Milliman states that: "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."

C. **CMS' claim that a FFS adjuster would result in inequity among plans is not credible or reasonable**

CMS provides an alternative rationale for not applying a FFS adjuster: even if the CMS technical study might otherwise support a FFS adjuster, the agency believes a FFS adjuster should not be applied because it would create inequities between audited and unaudited plans. We strongly disagree.

The FFS adjuster is needed to ensure compliance with the actuarial equivalence requirement in the statute. The issue of actuarial equivalence arises whenever CMS seeks to apply a documentation standard for payment (medical records) that differs from the standard used in developing the risk adjustment model (claims). Plans not subject to a RADV audit may still be subjected to different document standards if they face overpayment claims by the government, and two district courts have recently held that the actuarial equivalence requirement must be satisfied in the overpayment context. Thus, equity requires the application of a FFS adjuster to RADV audits. Further, unaudited plans that do not face payment recovery issues are not adversely affected by different documentation standards, and therefore the use of a FFS adjuster would not adversely affect them.

In addition, even assuming CMS was otherwise correct in identifying certain cases where there is a potential for different treatment, it would still be unreasonable for CMS to use the alternative rationale to avoid implementing a FFS adjuster. CMS seeks to justify not applying corrective action when such action is required to maintain actuarial equivalence for payments made to a specific MA plan undergoing a RADV audit, simply because no such corrective action will be taken by CMS with respect to MA plans not undergoing RADV audits. In other words, CMS wants to withhold fairness for some because the agency refuses to do justice to all. That interpretation is clearly not permitted under section 1853(a)(1)(C)(i) of the SSA. See *Global Tel*Link v. FCC*, 866 F.3d 397, 418 (D.C. Cir. 2017) (Silberman, J., concurring) ("*Chevron's* second step can and should be a meaningful limitation on the ability of administrative agencies to exploit statutory ambiguities, assert farfetched interpretations, and usurp undelegated policymaking discretion."). Moreover, the recent decision in *Poehling* confirms that the statute requires that CMS adjust payments to ensure actuarial equivalence. See *United States ex rel. Benjamin Poehling v. UnitedHealth Group, Inc.*, No. 16-08697, 2019 WL 2353125 (C.D. Cal. Mar. 28, 2019). CMS does not have the discretion to ignore that statutory requirement simply because it believes that requirement to be inequitable.

As a membership organization representing the MA industry, we can say without hesitation that our members support a FFS adjuster regardless of whether one of their contracts is selected for audit. We believe strongly that CMS is required to implement a FFS adjuster in payment recovery activities, and that such an adjustment is not only necessary to achieve actuarial equivalence but is equitable for both audited and unaudited plans in that context.

D. The FFS adjuster is required under the "same methodology" language in the statute

Section 1853(b)(4)(D) of the SSA requires that in computing expenditures for traditional Medicare, CMS must use the "same methodology as is expected to be applied in making payments to [MA plans]." CMS violates this statutory command if the "'methodology' applied in 'making payments' to [MA] insurers involves reconciliation based strictly on audited diagnosis codes for [MA] patients, in sharp contrast to unverified diagnosis codes for traditional Medicare patients from which payment rates were set." *UnitedHealthcare*, 330 F. Supp. 3d at 187. Just like the 2014 Overpayment Rule vacated in *UnitedHealthcare*, the Proposed Rule "fails to recognize a crucial data mismatch, and without correction, it fails to satisfy [1853(b)(4)(D)]." *Id.* A similar statutory interpretation – that this section of the statute is applicable to risk adjustment payments, and not just to CMS' annual reporting requirements as the Government had argued – was also recently upheld in the U.S. District Court for the Central District of California in its ruling in *United States ex rel. Benjamin Poehling v. UnitedHealth Group, Inc.:* "But the Court is unpersuaded that the statute is so limited [to CMS' annual reporting requirement], given that the face of the statute also requires '***computation*** [of] … [t]he average risk factor for the covered population . . . using the same methodology ***as is expected to be applied in making payments***' to MA plans." No. 16-08697, 2019 WL 2353125 at 5 (C.D. Cal. Mar. 28, 2019) (citing 42 U.S.C. § 1395w-23(b)(4)) (emphasis in original).

III. CMS' extrapolation proposal is procedurally defective, exceeds the Agency's statutory authority, and is arbitrary and capricious

A. The Proposed Rule fails to adequately reference the legal authority for extrapolation

CMS asserts that it may use contract-level extrapolation "based on longstanding case law and best practices from [the Department of Health and Human Services] and other federal agencies" (Preamble p. 54984). However, agency authority must be derived from statute, and the Proposed Rule never specifies what statute CMS believes grants it the authority to use extrapolation with respect to MA plans. As a result, the Proposed Rule violates the fundamental requirement imposed by the Administrative Procedures Act (APA) whereby a notice of proposed rulemaking must include "reference to the legal authority under which the rule is proposed."[29] The Proposed Rule also does not identify the "longstanding case law" referred to by the agency, thereby requiring the public to speculate regarding the decision(s) on which CMS might be relying. The APA's notice-and-comment requirements do not permit such an approach.

---

[29] 5 U.S.C. § 553(b)(2); *see also Attorney General's Manual on the Administrative Procedure Act* 29 (1947) ("The reference [to legal authority required by § 553(b)(2)] must be sufficiently precise to apprise interested persons of the agency's legal authority to issue the proposed rule.").

B. CMS does not have statutory authority to extrapolate RADV audit results

CMS does not have authority to use contract-level extrapolation against MA plans under the SSA. Most case law related to extrapolation does not address the threshold question of whether a federal agency has statutory authority to use extrapolation. Instead, it addresses the separate question of whether the use of extrapolation violates the constitutional right to due process.[30] The only appellate decision of which we are aware that addressed a somewhat similar statutory-authority question did so solely with respect to the use of extrapolation in FFS Medicare at a time when such extrapolation had already become a "long-standing and well-established practice" as applied to providers of services and suppliers under Medicare Parts A and B. *Chaves Cnty. Home Health Serv., Inc. v. Sullivan*, 931 F.2d 914, 923 (D.C. Cir. 1991). Even in that instance, however, the D.C. Circuit openly acknowledged that the question of statutory authority to use extrapolation was "close." *Id.* at 923. The D.C. Circuit also found that nothing in the Medicare Act at the time spoke directly to the use of extrapolation. *See id.* at 916–18. However, after repeatedly noting that the appellants (three home health agencies) failed to challenge the statistical validity of the calculations at issue, the court found that the use of extrapolation in the particular context before it represented a reasonable interpretation of the "authority to recoup overpayments *from providers*," *Id.* at 916-17, 921-22 (emphasis added).

Yet much has changed since the D.C. Circuit's decision in *Chaves County*, the statutory-authority holding of which has essentially gone untested in any other circuit court of appeals. Not only has that holding been undermined by the Supreme Court's rejection of the "novel project" of "Trial by Formula,"[31] the government itself has acknowledged the need for legislation before proceeding as suggested in the Proposed Rule, explaining in testimony before Congress:

> The President's Budget includes seven legislative and administrative proposals that will strengthen efforts to fight Medicare and Medicaid fraud and abuse . . .
>
> *Legislative Proposals Included in the Budget*
>
> Extrapolate MA Plan Sample Error Rate to Entire Plan Payment in Risk Adjustment Audits: Historically, CMS has only recovered overpayments from risk adjustment

---

[30] *See, e.g.*, *Ratanasen v. Cal. Dep't of Health & Human Servs.*, 11 F.3d 1467, 1469–71 (9th Cir. 1993) (addressing bankruptcy court's use of extrapolation with respect to amounts owed to state Medicaid program); *Yorktown Med. Lab., Inc. v. Perales*, 948 F.2d 84, 89–90 (2d Cir. 1991) (addressing state Medicaid agency's use of extrapolation); *Ill. Physicians Union v. Miller*, 675 F.2d 151, 154–56 (7th Cir. 1982) (same); *see also Mich. Dep't of Educ. v. U.S. Dep't of Educ.*, 875 F.2d 1196, 1204–06 (6th Cir. 1989) (addressing whether federal agency's use of extrapolation in recouping vocational-rehabilitation funds from State satisfied substantial-evidence standard); *Georgia ex rel. Dep't of Human Res. v. Califano*, 446 F. Supp. 404, 409–10 (N.D. Ga. 1977) (addressing whether federal agency's use of extrapolation in recouping Medicaid funds from State was arbitrary and capricious).

[31] *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 367 (2011). Writing for the Supreme Court in *Dukes*, Justice Scalia found that it was improper to certify a class action on the premise that the defendant would only be able to litigate its defenses with respect to monetary claims asserted by a sample of class members, the outcome of which would then be extrapolated to the class as a whole. *See id.* The Supreme Court recently went further by limiting the use of extrapolation to those instances where statistical evidence would be relevant in adjudicating an individual claim of liability. *See Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1046 (2016).

errors found in the audited sample. This proposal would require that CMS recover risk adjustment overpayments by extrapolating sample error rates to all audited plans through risk adjustment validation (RADV) audits. The plan payment will only be adjusted on a statistically valid sample of beneficiaries . . .[32]

It would "strain[] credulity to suggest that" the government submitted such a request to Congress "without analyzing the relevant statutes." *U.S. House of Representatives v. Burwell*, 185 F. Supp. 3d 165, 186 (D.D.C. 2016).

Further, Congress has not stood silent with respect to the use of extrapolation. Instead, it has authorized CMS to use extrapolation, but only with respect to a limited universe of Medicare overpayments and only under carefully prescribed circumstances.

In 2003, Congress added section 1893(f) to the SSA, entitled "RECOVERY OF OVERPAYMENTS." The new subsection (f) combined together a collection of overpayment-related provisions specific to a "provider of services or supplier," which are terms of art that refer to physicians, hospitals, and other entities but do not include MA organizations. The list of overpayment-related provisions for a "provider of services or supplier" included the use of repayment plans; limitations on recoupment; the provision of supporting documentation; the use of consent settlements; notice of code overutilization; and payment audits.

Importantly, subsection (f)(3), entitled "LIMITATION ON USE OF EXTRAPOLATION," states:

A [M]edicare contractor may not use extrapolation to determine overpayment amounts to be recovered by recoupment, offset, or otherwise unless [CMS] determines that—

(A)   there is a sustained or high level of payment error; or

(B)   documented educational intervention has failed to correct the payment error.

There shall be no administrative or judicial review under section 1869 [referring to appeal rights specific to Medicare Parts A and B], section 1878 [referring to additional appeal rights specific to certain providers of services under Part A], or otherwise, of determinations by [CMS] of sustained or high levels of payment errors under this paragraph.

The language of paragraph (3), which is included in the midst of a subsection focused entirely on overpayment issues related to providers and suppliers under Medicare Parts A and B, does not provide CMS with authority to use extrapolation with respect to anyone *other than* providers and suppliers under Medicare Parts A and B. "Statutory language cannot be construed in a vacuum. It

---

[32] *Departments of Labor, Health and Human Services, Education, and Related Agencies Appropriations for 2011: Hearings Before the H.R. Comm. on Appropriations*, 111th Cong. pt. 7 at 14 (2010) (written statement of William Corr, Dep't Sec'y, Dep't of Health & Human Servs.); *see also* Ctrs. for Medicare & Medicaid Servs., Dep't of Health & Human Servs., *Fiscal Year 2011 Performance Budget* 177 (2010) (describing proposal that would "[c]larify in statute that CMS can extrapolate the error rate found in the risk adjustment validation (RADV) audits to the entire MA plan payment for a given year when recouping overpayments").

is a fundamental canon of statutory construction that the words of a statute must be read in their context and with a view to their place in the overall statutory scheme." *Sturgeon v. Frost*, 136 S. Ct. 1061, 1070 (2016) (internal quotation marks and citation omitted). Nor can CMS use Congress's specific grant of extrapolation authority with respect to Medicare Parts A and B as an implicit grant of such authority with respect to Medicare Part C. *See, e.g.*, *Ry. Lab. Executives' Ass'n v. Nat'l Mediation Bd.*, 29 F.3d 655, 670 (D.C. Cir. 1994) (en banc) ("Unable to link its assertion of authority to any statutory provision, the [agency's] position in this case amounts to the bare suggestion that it possesses *plenary* authority to act within a given area simply because Congress has endowed it with some authority to act in that area. We categorically reject that suggestion.").

Furthermore, even if one were to view section 1893(f)(3) in isolation, the Proposed Rule makes no effort to explain how the statute's prerequisites for the use of extrapolation—*i.e.,* a determination of a "sustained or high level of payment error" or "documented educational intervention [that] has failed to correct the payment error"—have been satisfied with respect to those MA plans selected to undergo RADV audits. *See also* H.R. Rep. No. 108-391, at 785 (2003) (Conf. Rep.) (explaining that "[e]xtrapolation is limited to those circumstances where there is a sustained or high level of payment error, as defined by [CMS] in regulation, or document[ed] educational intervention has failed to correct the payment error").

C. CMS extrapolation proposal creates an unlawful presumption

Even if CMS has statutory authority to use extrapolation under certain circumstances with respect to MA plans, it would not mean CMS has statutory authority to use extrapolation in any manner that it chooses. For example, decisions such as *Chaves County* speak to the use of extrapolation on a provider-specific basis, where it is at least plausible that a provider who is demonstrated to have submitted incorrect payment claims under certain circumstances with respect to certain patients did so with respect to other, similarly situated patients under the same provider's care during the same time period.

The Proposed Rule, in contrast, would establish a system whereby CMS reviews a sample of patients treated by certain healthcare providers under contract with an MA plan, determines whether those particular providers maintained (in CMS' view) adequate medical documentation to support the diagnoses they reported to the MA plan, and applies an error rate to a universe of diagnoses reported to the MA plan by thousands of other, unrelated providers simply because they, too, are under contract with the MA plan. In doing so, CMS would essentially establish a presumption that is impossible for an MA plan to rebut. The MA plan would have no way of establishing the existence of sufficient supporting medical documentation related to the extrapolated universe of cases because the total overpayment amount will not be tied to specific providers and specific patients.

The establishment of such a presumption exceeds CMS' statutory authority. "[A]n agency is not free to ignore statutory language by creating a presumption on grounds of policy to avoid the necessity for finding that which the legislature requires to be found." *United Scenic Artists v. NLRB,* 762 F.2d 1027, 1034 (D.C. Cir. 1985). The creation of such a presumption "is beyond the

[agency's] statutory authority." *Id.* at 1035. At a minimum, there must be a "sound factual connection . . . between the facts giving rise to the presumption and the facts then presumed." *Holland Livestock Ranch v. United States,* 714 F.2d 90, 92 (9th Cir. 1983) (internal quotation marks and citation omitted).

No such connection exists with respect to the presumption created by the extrapolation regime contained in the Proposed Rule, which avoids using the word "presumption" even though CMS has previously acknowledged that the use of statistical sampling and extrapolation does, in fact, create a presumption. *See, e.g.,* Use of Statistical Sampling to Project Overpayments to Medicare Providers and Suppliers, HCFA Ruling No. 86-1, at 11 (Feb. 20, 1986) ("Sampling only creates a presumption of validity as to the amount of an overpayment which may be used as the basis for recoupment."). Just because a contracted provider maintained documentation that CMS, years after the fact, believes is insufficient to support diagnoses reported for a particular patient, does not logically suggest that the same is true of all patients treated by that provider, let alone that the same is true of all other providers under contract with the MA plan treating other patients. "The conclusion . . . simply does not follow from the premise," rendering the presumption beyond CMS' statutory authority. *United Scenic Artists,* 762 F.2d at 1035.

IV. CMS' published extrapolation methodology is so flawed that implementation would be arbitrary and capricious

AHIP commissioned a study by Wakely Consulting Group that examined the RADV sampling and extrapolation methodology applied in the contract-level audits conducted by CMS, but not finalized, for payment years 2011 to 2013.[33]

The Wakely study identified several significant areas of concern including:

- CMS' extrapolation approach is subject to a high degree of randomness and could result in inequitable treatment of similar contracts, because the application of the RADV process to contracts with similar average error rates may yield materially different payment penalties. The use of relatively small samples (201 enrollees), as well as the fact that coding errors can be rare, may result in erratic penalty results. To examine this issue, Wakely explored various scenarios of contract size and assumed coding error rates. For each scenario, Wakely ran 100,000 simulations of the RADV sampling process (selecting 201 enrollees). In one of the scenarios, Wakely assumed that 10 percent of HCCs are unsupported (coded but not supported by medical record), while 6 percent are supported but not reported (e.g., not coded but supported in medical record).[34] Wakely assumes a standardized bid of $850 per member per month (PMPM) for their simulations. They find wide variation in the penalties – from $0 PMPM to $67.07 PMPM – and note that "such variation in payment

---

[33] Murray, T., Morgan, E., Sauter, M. Medicare RADV: Review of CMS sampling and extrapolation methodology. Wakely Consulting Group. July 2018. Available at: https://www.ahip.org/wp-content/uploads/2018/07/Wakely-Medicare-RADV-Report-2018.07.pdf.

[34] Wakely also ran other scenarios based on different assumptions about the level of unsupported versus supported diagnoses – the assumption here is primarily used for illustrative purposes.

penalty for randomly chosen RADV samples from the same contract is obviously problematic."[35]

- The methodology is sensitive to which beneficiaries and conditions are included in the sample, because certain diseases can have a disproportionate impact on the payment error. As one example, based on a simulation of the RADV process, a single unsupported diagnosis of metastatic cancer could increase the payment penalty by 16.7 percent. Because the methodology does not take into account diagnosis-specific error rates, which are acknowledged by CMS to vary, penalties from one sample may be higher due to the "luck of the draw" for which diagnoses are selected (i.e., two different samples from the same plan could yield different payment penalties due to randomly selected diagnoses having a higher incidence of coding errors in the industry). Wakely notes that "random chance could drive material swings in extrapolated payment penalties" as a result.[36]

- The methodology could drive bias against higher enrollment contracts and contracts with low absolute risk scores. The sampling approach makes proportionally higher penalties more likely for larger enrollment contracts compared to smaller contracts. Wakely finds that "larger contract sizes are generally penalized by greater randomness in penalties."[37]

- The Proposed Rule explains that in choosing the enrollee population from which a sample will be taken for each MA contract selected for a RADV audit, CMS requires that enrollees have had "at least one diagnosis during the data collection year leading to at least one CMS-HCC assignment in the payment year." In other words, CMS excludes enrollees with no HCCs, thereby eliminating the possibility that such enrollees will be included in the sample from which CMS derives an overall payment error rate. Wakely notes that this practice "biases the sample payment error rate upwards." The report further explains: "Excluding non-HCC members from the RADV audit samples biases the observed payment error by removing potential supported but not reported codes for non-HCC members. This makes the expected observed payment error rate higher than the true payment error rate over the entire contract (RADV-eligible plus non-eligible)." In addition, the process for RADV that CMS uses for the exchange plans specifically considers the no-HCC population as its own stratum.[38]

- The methodology has a nonzero probability of yielding a payment penalty higher than the actual payment error, which is problematic even at very low probabilities. Such a penalty could yield a significant forfeiture of funding due to the randomness of the sampling methodology, and not due to coding accuracy.

---

[35] Ibid., p. 11.

[36] Ibid., p. 12.

[37] Ibid.

[38] Centers for Medicare & Medicaid Services, Center for Consumer Information and Oversight. 2017 benefit year protocols: PPACA HHS risk adjustment data validation [Version 2.0]. August 10 2018. Available at: https://www.regtap.info/uploads/library/HRADV_2017Protocols_Updates_v2.0_081018_v1_5CR_081018.pdf. For example, CMS notes on page 28 of this document "With the No-HCC population, the risk score errors will likely be under-statements, meaning the No-HCC risk scores should be adjusted upward."… "Consequently, there is some risk that CMS may be understating the error rate, variance, and risk score assumptions for the No-HCC stratum."

Therefore, even assuming for the sake of argument that CMS has statutory authority to extrapolate RADV audit results, AHIP believes the extrapolation methodology is so flawed that, if it were finalized, it would amount to an arbitrary and capricious agency action. For example, the methodological flaws leading to random results that cause similarly situated MA plans to be treated differently is a hallmark of arbitrary agency action. *See, e.g.*, *Cnty. of L.A. v. Shalala*, 192 F.3d 1005, 1022 (D.C. Cir. 1999) ("A long line of precedent has established that an agency action is arbitrary when the agency offers insufficient reasons for treating similar situations differently.") (internal quotation marks, citations, and brackets omitted). Further, as noted in Section III.C above, CMS presumes that provider documentation practices are uniform throughout the universe of providers under contract with an MA plan, a presumption that appears clearly arbitrary without evidence, particularly given that, to our knowledge, CMS has never applied extrapolation on anything other than a provider- or supplier-specific basis throughout the history of Medicare Parts A and B.

V.   CMS' retroactive application of the regulation is impermissible and not necessary or justified

Section 1871(e)(1) of the SSA specifies that a "substantive change in regulations, manual instructions, interpretative rules, statements of policy, or guidelines of general applicability under" title XVIII of the SSA shall not be applied retroactively to "items and services furnished before the effective date of the change" unless CMS makes one of two determinations. Either "such retroactive application is necessary to comply with statutory requirements;" or "failure to apply the change retroactively would be contrary to the public interest."

The Proposed Rule states that CMS intends to extrapolate RADV audit results beginning with payment year 2011. The Proposed Rule also states that even though the 2012 RADV Notice promised that CMS would apply the FFS adjuster in RADV audits beginning with those related to payment year 2011, CMS intends to break its promise with respect to past payment years. CMS solicits comment on whether applying the methodology to previous plan year audits would constitute retroactive rulemaking. However, CMS also indicates that even if doing so would constitute retroactive rulemaking, CMS will invoke authority under section 1871(e)(1)(A) to engage in such rulemaking.

The changes contained in the Proposed Rule clearly constitute retroactive rulemaking. We also believe the changes clearly exceed CMS' limited authority to engage in retroactive rulemaking.

A.   The Proposed Rule would constitute retroactive rulemaking

"To determine whether a rule is impermissibly retroactive, [a court] first look[s] to see whether it effects a substantive change from the agency's prior regulation *or* practice." *Ne. Hosp. Corp. v. Sebelius*, 657 F.3d 1, 14 (D.C. Cir. 2011) (emphasis added) (internal quotation marks and citation omitted). The Proposed Rule does both.

The Proposed Rule significantly revises the regulations that govern MA plans. For example, CMS would amend 42 C.F.R. § 422.310(e) by adding new language, stating: "MA organizations must

remit improper payments based on RADV audits and established in accordance with stated methodology, in a manner specified by CMS. For RADV audits, CMS may extrapolate RADV Contract-Level audit findings to Payment Year 2011 forward." Similarly, CMS would amend 42 C.F.R. § 422.311 by adding the following language: "Recovery of improper payments from MA organizations will be conducted according to the Secretary's payment error extrapolation and recovery methodologies. CMS will apply extrapolation to plan year audits for payment year 2011 forward."

The 2012 RADV Notice promised that CMS "would apply a FFS Adjuster as an offset before finalizing the audit recovery." CMS attempts to use that notice as evidence that implementing the Proposed Rule "would not upset any settled interest" as it relates to the use of extrapolation generally (Preamble p. 55040). However, this is clearly incorrect given the reversal on the FFS adjuster.[39] Further, existing case law demonstrates that the 2012 RADV Notice cannot be used to thwart a claim of retroactivity. *See, e.g.*, *Bowen v. Georgetown Univ. Hosp.*, 488 U.S. 204, 215 (1988) (finding rule change impermissibly retroactive even though it had first been announced in a notice published years earlier in the *Federal Register*); *Nat'l Mining Ass'n v. Dep't of Lab.*, 292 F.3d 849, 868 (D.C. Cir. 2002) (rejecting agency's argument against retroactivity where past agency practice was "encapsulated only in a manual, not in a regulation promulgated pursuant to notice-and-comment rulemaking").

We note that retroactivity in this context is not limited to RADV audits already undertaken for plan years 2011-2013; it covers any periods before the final rule is implemented and includes audits for 2014 that CMS recently initiated. In other words, CMS can only apply changes in RADV methodology to payment years after publication of a final rule, and plans must have the ability to factor the RADV rules into their bids. Thus, even if CMS were to finalize a proposal on extrapolation in MA in 2019, the earliest it could apply would be CY 2021.

B. <u>Retroactive application is not necessary to satisfy a statutory requirement</u>

The Proposed Rule asserts in passing that in retroactively applying the proposed changes, "CMS would be acting in compliance with" the Improper Payments Elimination and Recovery Improvement Act of 2012 (IPERIA). However, choosing a course of action that the agency (mistakenly) believes to be "in compliance with" a particular statute is fundamentally different from the determination required by section 1871(e)(1)(A)(i): namely, a determination that "such retroactive application *is necessary* to comply with statutory requirements." CMS made no such

---

[39] Separate from the general question of authority for retroactive rulemaking, we believe a refusal to honor the promise of a FFS adjuster would be arbitrary and capricious. In explaining a changed position, an agency must be "cognizant that longstanding policies may have 'engendered serious reliance interests that must be taken into account.'" *Encino Motorcars, LLC v. Navarro*, 136 S. Ct. 2117, 2126 (2016) (quoting *FCC v. Fox Television Stations, Inc.*, 556 U.S. 502, 515 (2009)). The Supreme Court has specifically cautioned agencies that "[i]t would be arbitrary and capricious to ignore such matters." *Fox Television*, 556 U.S. at 515. The entire MA community has reasonably relied on the 2012 RADV Notice, which CMS claimed at the time was a product of the agency "carefully review[ing] the more than 500 comments received on the draft methodology" that CMS published on its website in late 2010.

necessity determination in the Proposed Rule. Moreover, nothing in IPERIA *requires* CMS to engage in retroactive rulemaking in this context.

C. Retroactive application is not justified as being in the public interest

A public-interest determination under section 1871(e)(1)(A)(ii) is, at a minimum, subject to review under the arbitrary-and-capricious standard of the APA. *See, e.g.*, Sec'y Br. at 43, *St. Francis Med. Ctr. v. Azar*, 894 F.3d 290 (D.C. Cir. 2018). The APA, in turn, requires that an agency "examine the relevant data and articulate a satisfactory explanation for its action including a rational connection between the facts found and the choice made." *State Farm*, 463 U.S. at 43 (internal quotation marks and citation omitted). "In reviewing that explanation, [a court] must consider whether the decision was based on a consideration of the relevant factors and whether there has been a clear error of judgment." *Id.* (internal quotation marks and citations omitted). "Normally, an agency rule would be arbitrary and capricious if the agency has relied on factors which Congress has not intended it to consider, entirely failed to consider an important aspect of the problem, offered an explanation for its decision that runs counter to the evidence before the agency, or is so implausible that it could not be ascribed to a difference in view or the product of agency expertise." *Id.*

The public-interest determination is not justified in this case for several reasons.

First, if CMS could simply claim financial recovery as a basis, as it does here, CMS effectively would have almost limitless authority to implement changes retroactively. Essentially the public-interest exception would swallow the general rule against retroactive rules. And this interpretation would not be limited to RADV in the MA program; it in theory could apply to any one of the payment systems governing the traditional Medicare program.

Second, "agencies do *not* have free rein to use inaccurate data." *Dist. Hosp. Partners, L.P. v. Burwell*, 786 F.3d 46, 56 (D.C. Cir. 2015). As the D.C. Circuit recently emphasized in a case involving CMS, an agency "is required to 'examine *the relevant data* and articulate a satisfactory explanation for its action including a rational connection between the facts found and the choice made.'" *Id.* at 56–57 (quoting *State Farm*, 463 U.S. at 43) (emphasis supplied by D.C. Circuit). "If an agency fails to examine the relevant data—which examination could reveal, *inter alia*, that the figures being used are erroneous—it has failed to comply with the APA." *Id.* at 57.

In this case, the public-interest determination is predicated on the assumption that extrapolation of RADV audit results in past payment years will result in the "recoupment of millions of dollars of public money improperly paid to private insurers." To arrive at these estimates, the Proposed Rule mischaracterizes the level of alleged MA improper payments. CMS asserts that MA plans have had "high levels of payment error in the Part C program" (Preamble p. 55039, footnote 27). CMS says the "amount of improper payments" identified under the MA program is $14.35 billion or 8.31 percent of total MA payments in FY 2017 (Preamble p. 55039).[40] However, this figure represents the "gross" improper payment rate, which is a combination of two payment error estimates: 1) 'overpayments' to MA plans, and 2) 'underpayments' to MA plans. An overpayment

---

[40] This figure comes from the annual National RADV audit, conducted in accordance with IPERIA.

is defined as an instance where a diagnosis code submitted for payment purposes was not supported by the beneficiary's medical record. An underpayment occurs when the medical record review identifies an additional diagnosis code that should have been submitted to CMS and used for payment. The gross improper payment rate represents the sum of overpayments and underpayments; the two numbers are not netted. Therefore, underpayments increase the gross improper payment rate to the same degree as overpayments.

Use of the gross improper payment rate vastly overstates the purported impact to the government of errors in the MA program. CMS' estimates show that underpayments for FY 2017 comprise 35 percent of the $14.35 billion estimate of improper payments.[41] And that level is increasing; in FY 2018, underpayments were 42 percent of total improper payments.[42] We also note CMS has consistently found that the MA program has a significantly lower net improper payment rate than the FFS Medicare program. Chart 1 below shows the difference in the FFS and MA program gross and net improper payment amounts over time.

Chart 1. Underpayments and Overpayments at a Percent of Improper Payments, FFS vs. MA (FY2012-2018)
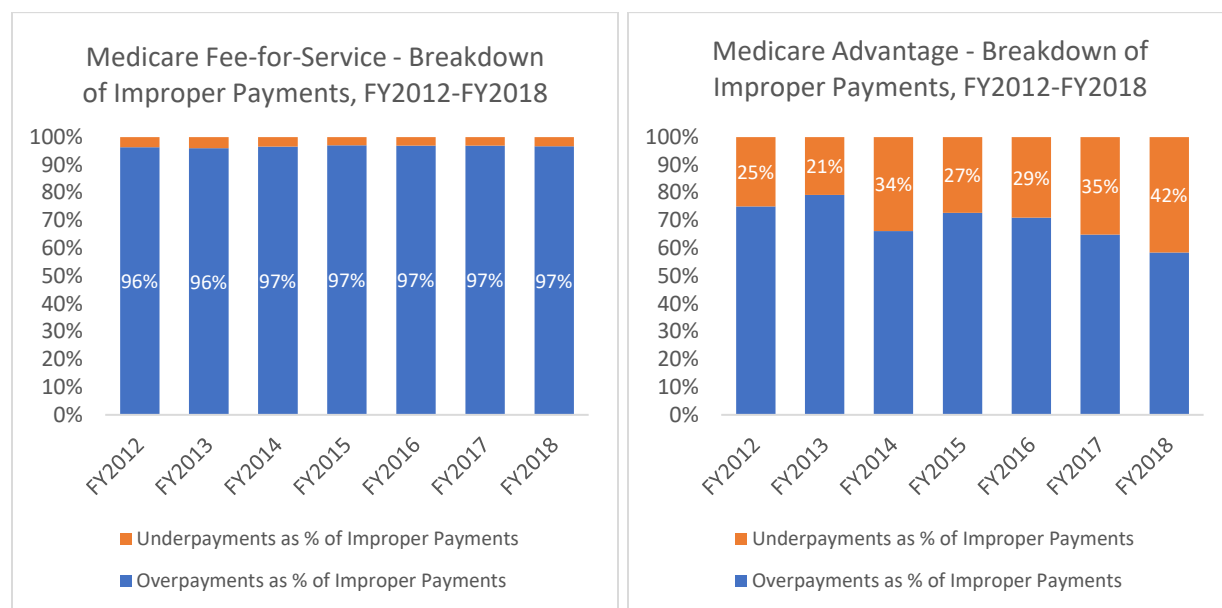
If the prevalence of underpayments in the MA program were properly considered, it would show a net improper payment rate of 1.37 percent or $2.6 billion in MA for FY 2018.[43] As shown in Chart 2 below, the MA net improper payment rate has decreased considerably since FY 2012, and in any event is much smaller than the FFS net improper payment rate. Accordingly, even if retroactive application of extrapolation were legally permissible in theory under the public-interest exception, we believe the relevant data demonstrates that the problem is small in relative
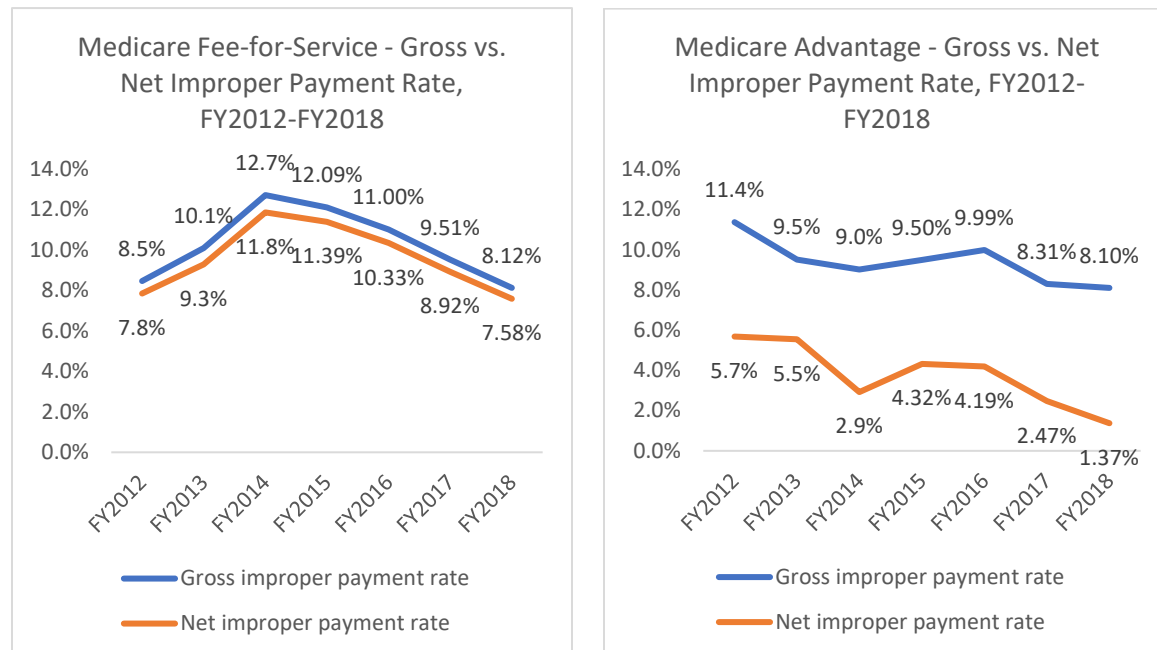
---

[41] Department of Health and Human Services. FY2017 Agency Financial Report. November 2017.
[42] Department of Health and Human Services. FY2018 Agency Financial Report. November 2018.
[43] Ibid.

terms, is shrinking, and therefore does not support a step as extraordinary as retroactive rulemaking.

Chart 2. Gross vs. Net Improper Payment Rates, FFS vs. MA (FY2012-2018)

Medicare Fee-for-Service - Gross vs. Net Improper Payment Rate, FY2012-FY2018

Gross improper payment rate: 8.5%, 10.1%, 12.7%, 12.09%, 11.00%, 9.51%, 8.12%
Net improper payment rate: 7.8%, 9.3%, 11.8%, 11.39%, 10.33%, 8.92%, 7.58%

Medicare Advantage - Gross vs. Net Improper Payment Rate, FY2012-FY2018

Gross improper payment rate: 11.4%, 9.5%, 9.0%, 9.50%, 9.99%, 8.31%, 8.10%
Net improper payment rate: 5.7%, 5.5%, 2.9%, 4.32%, 4.19%, 2.47%, 1.37%

Third, CMS' public-interest determination fails to consider the interest of finality, which is conspicuous for at least two reasons. CMS has long cited the interest of finality as a principal reason not to upset Medicare payment determinations. In addition, CMS has gone so far as to cite the interest of finality as a reason to engage in retroactive rulemaking under section 1871(e)(1)(A)(ii).[44]

The interest of finality serves more than just financial considerations. As CMS recently explained to the D.C. Circuit, the interest of finality also reflects "evidentiary and administrability considerations. Records grow stale, memories fade, personnel move on, and retention is costly."[45] This is especially true in the case of RADV audits. Such audits are predicated on the review of medical records related to services that may have been provided many years earlier.

These operational barriers to retroactive rulemaking are illustrated in the RADV audits for 2014 that CMS began earlier this year. Under these audits, plans must collect medical records from providers for services rendered in 2013. Because of the passage of time, medical records from 2013 may not exist or may be nearly impossible for plans to retrieve, for numerous reasons:

---

[44] *See* Hospital Outpatient Prospective Payment and Ambulatory Surgical Center Payment Systems, 78 Fed. Reg. 74,826, 75,165 (Dec. 10, 2013). In this rule, CMS asserted that not applying proposed regulatory changes retroactively would "undermine . . . the interests of both the Medicare program and Medicare providers in the finality of reimbursement determinations."

[45] Sec'y Br. at 47, *St. Francis Med. Ctr. v. Azar*, 894 F.3d 290 (D.C. Cir. 2018).

- Solo practitioners have passed away, with no accessible repository for old medical records.
- Providers switched electronic medical record systems and have IT challenges in producing medical records prior to the switch (particularly for patients no longer seen by the provider and whose records were never migrated to the new system).
- Providers placed their records into off-site storage facilities whose personnel cannot locate the records in a timely fashion.
- Mental health providers are unwilling to provide data due to privacy concerns (that is, these providers will not release medical records without explicit beneficiary permission that may be difficult or impossible to obtain).

In addition to increasing the risk that records cannot be found to substantiate diagnoses, the long delay in the 2014 plan year audit will contribute to other documentation challenges for diagnoses. For example, provider signatures may be incomplete on some medical records, but if the providers can no longer be located to complete the attestations, CMS will not accept the records. In addition, since October 1, 2015, providers have used ICD-10 codes. However, the HCCs for the 2014 audit were based on ICD-9 codes which are no longer in use. Moreover, two different CMS-HCC risk adjustment models were used for payment in 2014. Even assuming plans still employ or can hire medical record reviewers who are knowledgeable about these outdated provisions, the situation lends itself to more disputes between plans and CMS.

Thus, if CMS is allowed to apply extrapolation retroactively, the agency will artificially inflate the number of cases identified as coding errors and the resulting amount of alleged overpayments. Diagnoses may be found unsubstantiated, not because the beneficiaries' clinical conditions did not exist, but because their medical records could not be obtained or validated (e.g., because providers are deceased or retired). These impacts stem directly from CMS' long delay in initiating the 2014 audits.

Fourth, we are concerned that CMS further justifies its proposal by failing to acknowledge the clinical importance of accurate diagnosis coding in MA. CMS states that "there is an incentive for plans to potentially over-report diagnoses so that they can increase their payment" (Preamble p. 55037). This statement fails to acknowledge that identifying accurate diagnoses is a crucial mechanism for understanding the health status of a patient. Due to the nature of capitated payments, MA promotes accurate diagnosis coding to support coordinated and integrated care because plans must consider the entire patient and how each of their clinical conditions interact. More accurate and detailed diagnosis coding helps plans identify and support the specific health care needs of their enrollees to ensure they receive integrated care coordination and chronic disease management.

Fifth, we note it would not be in the public interest to invoke section 1871(e)(1)(A)(ii) to retroactively "fix" CMS' failure to engage in notice-and-comment rulemaking with respect to the 2012 RADV Notice (see Section VI below). *Cf. Georgetown Univ. Hosp. v. Bowen*, 821 F.2d 750, 758 (D.C. Cir. 1987) ("The Secretary's suggestion that retroactive rulemaking is permissible to remedy a procedural defect in a rule would, if accepted, make a mockery of the provisions of the APA. Obviously, agencies would be free to violate the rulemaking requirements of the APA with

impunity if, upon invalidation of a rule, they were free to 'reissue' that rule on a retroactive basis."), *aff'd on other grounds*, 488 U.S. 204 (1988).

VI. Implementation of the RADV audit methodology would violate rulemaking requirements

The Proposed Rule is procedurally invalid because it fails to comply with the notice-and-comment provisions applicable to substantive rules under the Administrative Procedures Act (APA). *See* Section 5 U.S.C. 553. CMS, in addition, is bound by the Medicare-specific rulemaking requirements in the Medicare Act. Section 1871(a)(2) of the SSA states that "[n]o rule, requirement, or other statement of policy (other than a national coverage determination) that establishes or changes a substantive legal standard governing . . . the payment for services . . . under this title shall take effect unless it is promulgated by the Secretary by regulation." The U.S. Supreme Court in *Azar v. Allina Health Services* recently upheld a D.C. Circuit Court opinion which invalidated how CMS calculated certain hospital payments under Medicare Part A because the methodology was not issued through regulation as required by section 1871(a)(2). 139 S. Ct. 1804 (2019). The Court noted that the policy "dramatically—and retroactively—reduced payments to hospitals serving low-income patients" and that "[b]ecause affected members of the public received no advance warning and no chance to comment first, and because the government has not identified a lawful excuse for neglecting its statutory notice-and-comment obligations" the agency's policy must be vacated. Id. at 1808.[46]

The Proposed Rule has components that are very similar to the policy invalidated in the *Allina* case. For example, the Proposed Rule applies extrapolation to 2011-2013 audits, a process that can result in "dramatically—and retroactively—reduced payments" to MA plans, using an extrapolation methodology never developed through rulemaking. CMS proposed its methodology for conducting RADV audits, including extrapolation, via the CMS website on December 20, 2010 instead of the Federal Register (Preamble p. 55038). The agency asked for comments 30 days after publication, and not 60 days, as is required under section 1871(b)(1). CMS then published the final methodology in the 2012 RADV Notice, which was essentially the same methodology that they had proposed, with one notable change – CMS acknowledged the need for the FFS adjuster, to adjust for different substantiation standards between MA payment and model development. Neither the 2010 proposal nor the 2012 RADV Notice had any discussion of alternatives considered, the impact on the industry, the rationale for the policy, or how the methodology fit with existing regulatory or statutory requirements.

The Proposed Rule attempts to justify the process for developing the 2012 RADV Notice, noting that "we invited public comment on this proposed methodology, and received more than 500 comments, which we carefully reviewed" (Preamble p. 55038). However, as summarized above,

---

[46] While the Court did not expressly endorse how the appeals court defined "substantive legal standard" under 1871(a)(2) or provide detailed guidance about how to interpret that phrase, the Court upheld the lower court because none of the Government's legislative history and policy-based arguments for avoiding the rulemaking requirement were persuasive. *Azar v. Allina Health Services*, 139 S. Ct. 1804, 1814-16 (2019).

that process failed to incorporate key elements of a formal rulemaking process. Moreover, while CMS may have carefully reviewed 500 comments, they did not respond substantially or directly to any of them, including comments raised by AHIP, except as related to one item: the FFS adjuster. This prevented the public from having an opportunity to learn why the agency decided to make policy as it did, and how the agency would respond to concerns of affected stakeholders.[47]

We recognize CMS is now requesting comments through the Proposed Rule on the audit methodology. However, we understand the comment request does not affect audits already conducted from 2011 to 2013, because it would be entirely impractical to conduct new audits on those years given the passage of time.

In addition, the Proposed Rule states that CMS will develop a new RADV methodology for audit years after 2013. We believe this new RADV methodology would also reflect a change in a substantive legal standard governing payments and therefore require rulemaking under *Allina*. However, CMS clearly does not intend to provide stakeholders with the opportunity to meaningfully analyze and provide comments on the proposed methodology for audit years after 2013. For example, the Preamble states that "CMS is not required to set forth the methodology for calculating an extrapolated payment error through regulatory provisions." (Preamble p. 55038). While CMS says that in the "interest of transparency", it would describe its intent to develop a new RADV methodology, CMS' description of the methodology in the Proposed Rule for audit years after 2013 is far too vague to meet rulemaking requirements.[48] In addition, the agency is actively conducting audits for 2014, using a new methodology, before the comment period on the rule closed. This shows the agency has pre-judged the issues raised in the Proposed Rule.

Moreover, despite initiating audits for 2014, CMS has yet to provide the type of technical details that stakeholders need to understand CMS' methodology. For example, CMS held a training on April 2, 2019 limited to only those contracts selected for the 2014 audit. CMS provided limited details on the new methodology, including a two-tiered approach to sampling that reflects a sub-cohort methodology.[49] No information in the training was proposed for public comment prior to

---

[47] *American College of Emergency Physicians v. Price*, 264 F.Supp.3d 89, 94 (D.D.C. 2017) ("Although an agency 'need not address every comment' made during the notice and comment period, 'it must respond in a reasoned manner to those that raise significant problems.'"(citations omitted))

[48] In a brief outline of the new methodology, CMS says that it "would calculate improper payments made on the audited MA contract for a particular sub-cohort or sub-cohorts in a given payment year." CMS further says that its methodology would be based on "statistically valid sampling and extrapolation methodologies." The agency also indicates that sub-cohorts could be "enrollees for whom a particular HCC or one of a related set of HCCs was reported." CMS "could often use a much smaller sample size" while generating "statistically significant recoveries." (Preamble p. 55039).

[49] In what CMS calls "Tier One", the sample would consist of 299 enrollees across 131 MA contracts. This cohort was based on the enrollees with "the highest predicted overpayment." CMS would not extrapolate results from Tier One audits. In what CMS calls "Tier Two", CMS would apply a sub-cohort methodology to enrollees with a high predicted overpayment rate – estimated through a regression model, the details for which have not been provided – and that have diabetes. Only 32 beneficiaries are being sampled per contract for this sub-cohort methodology, which is well below the sample size of 201 beneficiaries used in the 2011-2013 audits. CMS would extrapolate the results from Tier Two audits after the proposed regulation is finalized. Additionally, rather than auditing 30 contracts as CMS has historically selected (and suggested in its impact analysis of the Proposed Rule), CMS has selected 188 contracts for Tier Two audits.

being shared. This training has not been publicly posted on the CMS RADV website. Accordingly, the public has not had an opportunity to determine whether CMS' approach would be statistically valid or to otherwise adequately assess the proposal. Critical issues that are not addressed or left unclear include the following:

- The process for selecting sub-cohorts for audit purposes.
- The sizes needed to calculate a "statistically significant extrapolated recovery" or even what is meant by a "statistically significant extrapolated recovery."
- How CMS determines which contracts would be audited.
- How CMS would extrapolate the findings from these audits.

CMS has also not disclosed which contracts were selected for the 2014 RADV audit. However, CMS has a public document available on its website that lists every contract selected for RADV audit by year since 2007.[50] In addition to publicly releasing more detailed information about the audit methodology being used for the 2014 audit, we urge CMS to update this document with the list of contracts selected.

We also have serious concerns with the CMS statement in the Preamble that "we would make any future changes to that methodology (or those methodologies) through the Health Plan Management System." The agency is clearly stating an intent to use sub-regulatory guidance to issue RADV policy in the future, much as it did to establish the 2011 to 2013 audit methodology. We believe this is inconsistent with the SSA requirement for notice-and-comment rulemaking, as indicated in the *Allina* case. It also signals an unfortunate lack of CMS willingness to engage stakeholders on this critical issue.

VII.Other Issues

A. <u>**CMS's substantiation standards are insufficient for determining if the patient has the disease**</u>

On page 55037 of the Preamble of the Proposed Rule, CMS discusses the need for medical record documentation of diagnosis codes to support payment. The agency points to sub-regulatory guidance – none of which has ever been subject to notice-and-comment rulemaking – in which it has explained this requirement "since the beginning of the MA program." The only CMS RADV guidance that arguably satisfied the requirement was CMS' rule proposed in 2009 and finalized in 2010.[51,52]

Commenters to the 2009 rule expressed concern that the medical record requirement was overly proscriptive. It did not adequately consider the fundamental issue at stake – whether a person in

---

[50] Centers for Medicare & Medicaid Services. Medicare Advantage Risk Adjustment Data Validation Audits Fact Sheet (updated June 1, 2017). Available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/recovery-audit-program-parts-c-and-d/Other-Content-Types/RADV-Docs/RADV-Fact-Sheet-2013.pdf

[51] Medicare Program; Policy and Technical Changes to the Medicare Advantage and the Medicare Prescription Drug Benefit Programs; Proposed Rule, 74 Fed. Reg. 54633 (2009).

[52] Medicare Program; Policy and Technical Changes to the Medicare Advantage and the Medicare Prescription Drug Benefit Programs; Final Rule, 75 Fed. Reg. 19677 (2010).

question actually has a disease. CMS noted such concerns, stating that "commenters contended that the one best medical record policy forces plans to omit relevant data that could be supported through documentation that CMS does not permit – such as prescription drug data and lab results."

However, CMS did not propose any changes to address these concerns. They responded instead that "the RADV risk adjustment model is based upon FFS claims data from specific risk adjustment provider types, and not alternative data sources, such as, prescription drug data or lab results. Therefore, the RADV audit process is based upon supporting medical record documentation from provider data sources that are used to calibrate the model."

If CMS were to finalize its proposal on extrapolation, we believe it would need to revisit its position on documentation. The purpose of the RADV program should be to determine if a person's diagnosis is correct. Absence of documentation in a RADV audit does not mean a person does not have the disease or is not being treated for it. Rather, it simply means that a medical record could not be located that supports the diagnosis code.

CMS has never adequately explained its continued refusal to consider other sources of information that could substantiate a diagnosis code. Medical records themselves are indications of what the person is being actively treated for, and not necessarily of what conditions the person has. For example, a person who has diabetes under control through medication may see a provider for a different condition. The medical record in this situation may not reflect that this person has diabetes, yet a simple check of prescription data for that person would show the person is taking insulin and therefore does, in fact, have diabetes. The person may also be receiving other services related to the diabetes. Because the HCC is meant to capture the incremental costs of an individual with diabetes, excluding the diagnosis code here would be inaccurate. Yet that is what CMS would do in the RADV audit by virtue of depending only on the medical record for proof that the person has diabetes.

CMS also seems to be under the impression that diagnosis coding is precise. In reality, coding is not an "exact science" and reasonable people can interpret the same medical record in different ways. Coding guidelines can be unclear and interpreted differently by different people. And while CMS references the ICD-10-CM guidelines, it did not publish the guidelines that it was using for the 2011 to 2013 RADV audits. In this sense, plans were completely in the dark about what guidelines CMS would use for these audits.

For these reasons, CMS should revisit its requirement that HCCs can only be supported by the medical record and should consider alternative sources of data to substantiate diagnosis codes.

B. <u>Potential recovery based on OIG findings raises serious concerns</u>

We strongly oppose the suggestion in the Preamble that MA organizations could be forced to remit payments based on findings from OIG audits. As CMS indicates in footnote 25 on page 55039, OIG "does not seek comment on its methodology for risk adjustment audit work that may lead to overpayment recoveries from MA organizations." Although OIG is required by statute to follow generally accepted government auditing standards, this requirement does not adequately

account for the necessary actuarial and statistical methodological principles at issue in the complex system of MA risk adjustment and payment recovery. Several MA organizations raised similar concerns about the potential for lack of consistency in methodology and audit process between CMS and OIG when CMS proposed to expand authority for conducting RADV audits in 2014.[53] At that time, CMS finalized its proposal to specify that OIG has the authority to conduct RADV audits but did not address these concerns.

Because OIG does not seek comment on its methodology and is not required to employ the same methodology as finalized by CMS after formal notice-and-comment rulemaking, it is possible that CMS and OIG could arrive at different conclusions about the payment accuracy of the same MA contract due to different methodologies. For example, CMS released the "Contract-Level Risk Adjustment Validation Medical Record Reviewer Guidance"[54] that would be used to determine whether a medical record supported a given diagnosis for any RADV audit that occurs after September 2017; however, OIG has never released the diagnosis coding guidelines that it is currently using or will use in the future to establish supporting documentation for a diagnosis in a medical record. In addition, CMS does not provide any additional information on what remedies would be available to plans that do not concur with audit findings by the OIG. For example, if plans were forced to remit payments based on OIG audit findings then they should have the same right to appeal those findings as to appeal CMS audit findings, and CMS in conjunction with OIG would need to establish such as an appeals process.

Therefore, MA organizations could be subject to conflicting RADV audit methodologies employed by different government agencies with possibly divergent diagnosis coding guidelines that potentially could have different appeal rights. In fact, unless CMS and OIG coordinate appropriately, it is possible an MA organization could be subject to conflicting results for the same people. From the perspective of an MA organization, it would be entirely arbitrary which government agency chose one of its contracts to audit. Furthermore, the regulations governing RADV audits apply to all RADV audits and do not distinguish between RADV audits conducted by CMS and those conducted by OIG or any other government agency.

Unless OIG is required to conduct RADV audits using the exact same methodology employed as CMS, we recommend that CMS rescind this proposal.

C. Further detail is needed on potential CMS RADV appeals proposal

CMS indicates that the agency is considering whether to explicitly expand MA organizations' appeal rights in RADV. It describes one option as adopting practices for MA plans to appeal the RADV payment error calculation methodology similar to those for providers and suppliers in Medicare FFS Parts A and B.

---

[53] Medicare Program; Contract Year 2015 Policy and Technical Changes to the Medicare Advantage and the Medicare Prescription Drug Benefit Programs; Final Rule, 79 Fed. Reg. 29843 (2014).

[54] This document was first released on September 27, 2017 for audits commencing after that date. On March 20, 2019, CMS released an updated document in effect as of that date but applied retrospectively for audits commencing after September 27, 2019.

As CMS has not made a specific proposal to expand MA organization appeal rights under RADV audit, it is not clear how the agency is considering applying the Medicare FFS appeals framework to the RADV program. We therefore are unable to provide specific feedback in response to CMS' invitation for comments on this point.

In addition, the RADV audit appeals process in 42 C.F.R. § 422.311 unnecessarily restricts the rights of MA organizations by allowing them to appeal only one medical record per HCC. We urge CMS to expand the number of medical records that can be appealed for each HCC audited, to allow a more complete review of key evidence that can substantiate a clinical diagnosis.

D. RADV guidance related to encounter data is needed

CMS points out that plans submit diagnoses through two systems – the Risk Adjustment Processing System (RAPS) and the Encounter Data System (EDS) (Preamble p. 55037). However, these two systems are quite different. In RAPS, plans pre-identify all diagnoses to be submitted to CMS. In EDS, plans submit all claims data – also known as encounter data – and CMS then identifies which diagnoses should be used for risk adjustment using what is known as the "filtering logic".

Nowhere in the Preamble does CMS address the application of RADV in an encounter data setting. Plans have submitted encounter data to CMS since 2012, and CMS has selected diagnoses from these data to calculate payment since 2015. By 2022, CMS anticipates that EDS will be the only source for diagnoses used for payment (Part I of the Advance Notice for 2019).[55] The EDS rules are such that plans are required to submit all data to CMS – regardless of whether these data are to be used for risk adjustment. And then, CMS determines through its own set of rules which diagnoses are allowed for risk adjustment.

We believe CMS needs to give serious consideration to potential changes that may be required in RADV to reflect the EDS process. The agency should collaborate closely with industry to develop that approach and address this issue through future rulemaking.

VIII. Recommendations

Based on the foregoing:

- We urge the agency to withdraw the RADV Proposed Rule.
- We ask that the agency affirm that it cannot apply regulations retroactively.
- The agency should acknowledge that, in the absence of recalibrating the HCC model using audited FFS diagnosis data, a FFS adjuster is required under statute whenever it attempts to determine the accuracy of risk adjusted payments to MA plans by auditing MA diagnosis data against the medical records, and improve the audit methodology.

---

[55] Centers for Medicare & Medicaid Services. Advance Notice of Methodological Changes for Calendar Year (CY) 2019 for the Medicare Advantage (MA) CMS-HCC Risk Adjustment Model. December 27, 2017. Available at: https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/Advance2019Part1.pdf.

- We urge CMS to engage in meaningful, collaborative dialogue with the industry to develop RADV methodological changes going forward, and to ensure they are implemented solely through notice-and-comment rulemaking and on a prospective basis.

Medicare Advantage RADV FFS adjuster: White paper
Commissioned by America's Health Insurance Plans (AHIP)
August 23, 2019

Rob Pipich, FSA, MAAA

Milliman

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 "*calibration*." 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://dlbjbjzqnk95t.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 $0.5 \wedge 6 = 1.6\%$. 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 $0.5 \char`^ 2 = 0.25$ for the first beneficiary and $0.5 \char`^ 10 = 0.0001$ 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 $0.5 \char`^ 6 = 0.016$ 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 BENEFICIARIES	ON CLAIM?	FFS COST ACTUAL	FFS COST 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	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

FFS BENEFICIARIES	MA PAYMENT WITHOUT FFS ADJUTER			MA PAYMENT WITH FFS ADJUSTER		
	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
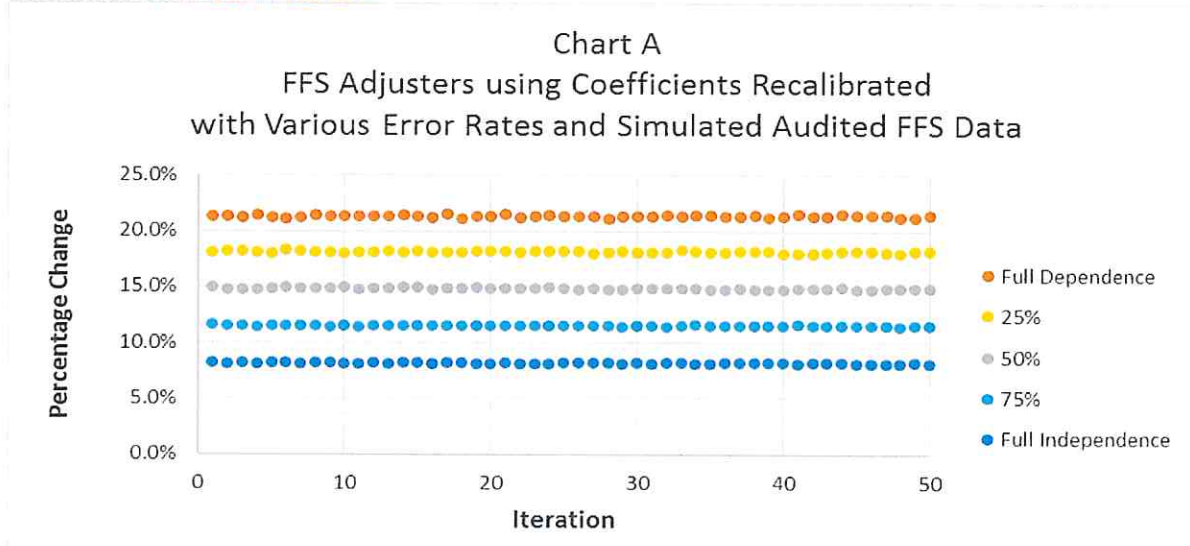
Under the midpoint error rate scenario and based upon 50 iterations, the FFS adjuster is between 14.85% and 14.90%, with a 99% level of confidence.

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

FFS BENEFICIARIES	MODEL CALIBRATED AND NORMALIZED WITH UNADJUSTED FFS DIAGNOSES				MODEL CALIBRATED AND NORMALIZED WITH ADJUSTED FFS DIAGNOSES			
	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

FFS BENEFICIARIES	ON CLAIM?	MODEL CALIBRATED WITH ADJUSTED FFS DIAGNOSES BUT NORMALIZED WITH UNADJUSTED DIAGNOSES		MA PAYMENT WITHOUT FFS ADJUSTER		MA PAYMENT WITH FFS ADJUSTER	
		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

FFS BENEFICIARIES	MODEL CALIBRATED AND NORMALIZED WITH ADJUSTED FFS DIAGNOSES		MA PAYMENT WITHOUT FFS ADJUSTER	
	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

FFS BENEFICIARIES	ON CLAIM?	MODEL CALIBRATED AND NORMALIZED WITH UNADJUSTED FFS DIAGNOSES: AFTER NORMALIZING	MA PAYMENT WITHOUT FFS ADJUSTER* BEFORE RADV	AFTER RADV	MA PAYMENT WITH FFS ADJUSTER* 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

Case 1:16-cv-00157-RMC   Document 44-4   Filed 10/02/17   Page 3 of 7

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.
-
-
-

(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.
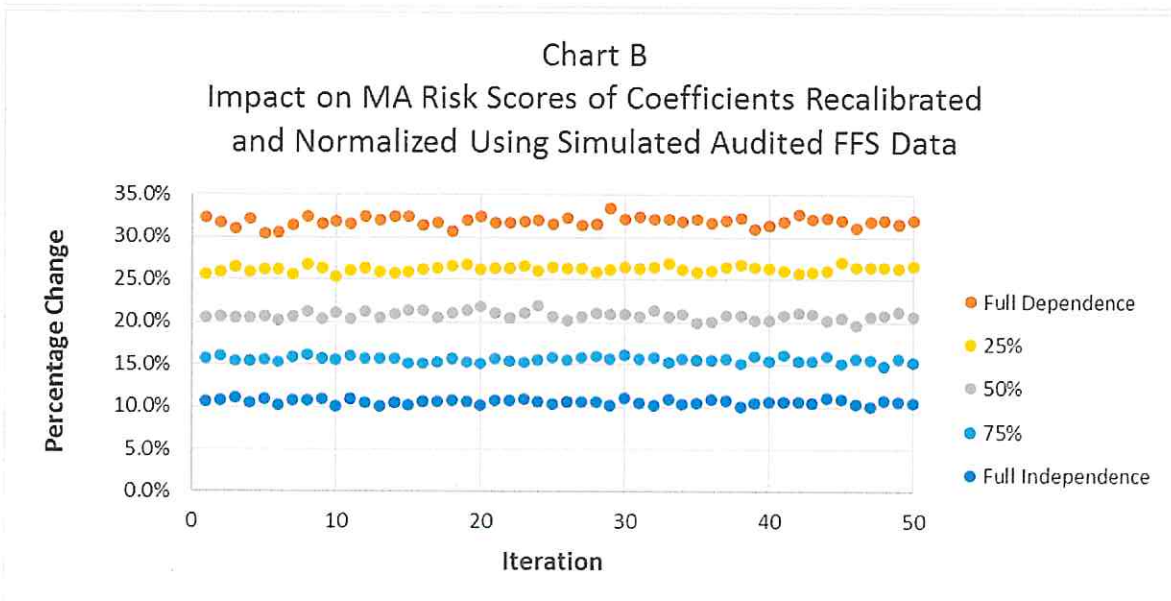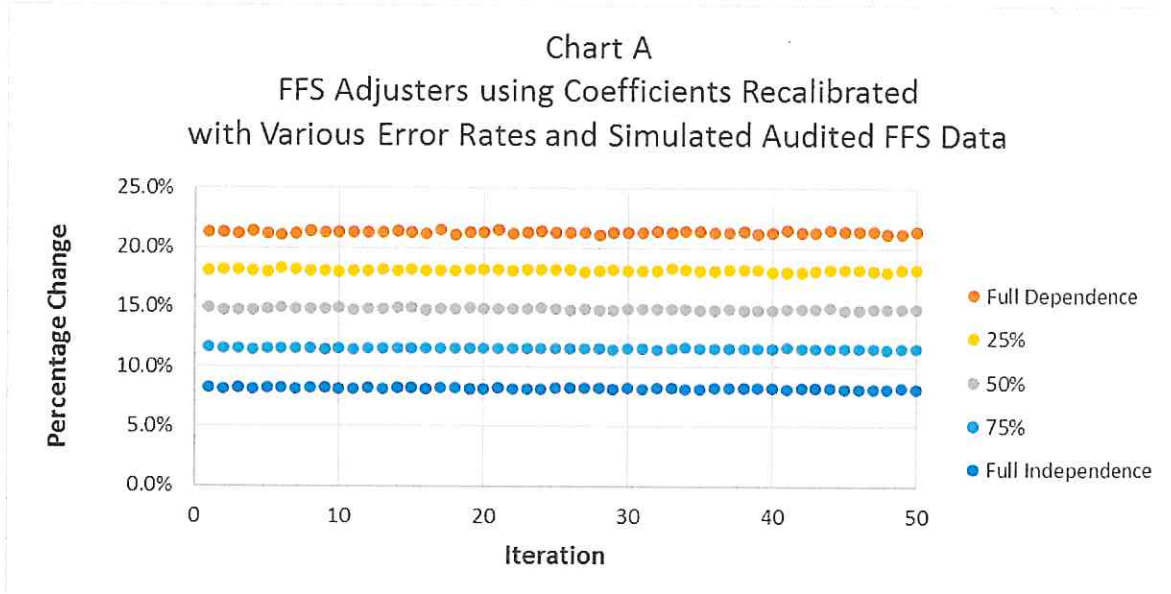
Appendix F: Statistical results from 50 simulations

Chart A
FFS Adjusters using Coefficients Recalibrated
with Various Error Rates and Simulated Audited FFS Data

Chart B
Impact on MA Risk Scores of Coefficients Recalibrated
and Normalized Using Simulated Audited FFS Data

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%

Wakely

Medicare RADV:
Review of CMS Sampling and Extrapolation Methodology

July 2018

Prepared by:
Wakely Consulting Group

Tim Murray, FSA, MAAA
Senior Consulting Actuary

Evan Morgan, ASA, MAAA, PhD
Consulting Actuary

Matt Sauter, ASA, MAAA
Consulting Actuary

Contents

Executive Summary

The Centers for Medicare & Medicaid Services (CMS) conducts Medicare Advantage (MA) Risk Adjustment Data Validation (RADV) audits as part of its program integrity efforts. Per CMS, MA RADV audits evaluate whether the diagnosis codes submitted by Medicare Advantage Organizations (MAOs), which directly influence CMS payments to MAOs, can be validated by supporting medical record documentation. In February of 2012 CMS published incomplete details of its RADV audit payment error calculation methodology, which comprises a sampling method and a payment error (penalty) extrapolation method. This report endeavors to provide an overview and technical evaluation of CMS methodologies, but not to provide a comprehensive evaluation of RADV audit process operations.

The CMS RADV audit methodology seeks not only to measure the payment error rate of selected MA contracts, but also to retroactively adjust CMS payments downward in instances where the CMS-derived payment error is higher than a to-be-defined coding accuracy standard. It is not administratively feasible for CMS to review the universe of medical record documentation for RADV contracts. Therefore, CMS must rely on a sampling method to approximate the payment error rate. CMS has indicated its intent to extrapolate the observed sample payment error across the MA contract's RADV-eligible population (oftentimes the large majority of the contract population). As a result, the CMS payment error extrapolation approach means that payment recoupment will affect not only revenue associated with MA beneficiaries whose medical records are audited, but also beneficiaries whose records are not audited.

Our technical evaluation explored how well the CMS sampling approach approximates the true payment error rate. The payment error rate reflects the combined impact of the coding error rate (frequency of coding errors) and the magnitude (risk score value or severity) of coding errors. The coding error rate reflects both the percentage of unsubstantiated codes and the percentage of supported but not submitted codes. The magnitude (severity) of coding errors is driven by the risk score value of specific coding errors, which may vary widely based on the morbidity profile (mix of diagnoses) of each MAO contract. Given the CMS-stated intent to extrapolate sample payment errors to retrospectively recoup MAO payments, we evaluated potential drivers of bias and inequity in the payment error extrapolation calculation. Specifically, we evaluated whether contract attributes other than the coding error frequency (e.g. contract size, diagnostic profile, average risk score) could potentially drive inequitable penalties. We also evaluated the risk that contracts with the same average payment error rate may experience inequitable payment penalties.

In order to perform a technical evaluation of CMS's RADV methodology, we simulated the RADV process on Medicare Limited Data Set (LDS) Standard Analytical Files. We used Monte Carlo simulation to replicate the CMS RADV process more than two million times on actual Medicare beneficiary claim and diagnosis data, varying MA contract sizes and assumed coding error rates.

As detailed in Appendix A, Monte Carlo simulation is a commonly-used mathematical technique to measure the statistical characteristics of processes that involve random variables and random sampling.

Our detailed analysis yielded a number of key findings and observations, summarized briefly below:

- The extrapolated payment error calculation is subject to a **high degree of randomness** - the sampling methodology utilized by CMS has the potential to accurately estimate the contract-specific payment error rate, but payment error extrapolation allows for MA contracts with identical coding error rates to pay vastly different penalties.

- The payment error calculation is **very sensitive to small variations in diagnostic mix** of the sample population - MA contracts with varying risk profiles by disease state may be subject to materially variant extrapolated penalties. We illustrate via simulation that even a single coding error in the randomly chosen RADV sample may drastically impact the RADV payment error penalty result.

- The CMS methodology tends to levy **disproportionate payment error penalties** on higher enrollment contracts and low absolute risk score contracts.

- CMS's MA RADV approach gives no consideration to **diagnosis-specific substantiation rates** – an MA contract may have a high prevalence of hard-to-substantiate diagnosis codes and therefore a high expected coding error rate. CMS does not adjust its penalty calculation to account for this dynamic despite publicly acknowledging the potential for such diagnosis-specific variation in several recent regulatory publications.

- Extrapolated payment penalties have the potential to be materially **larger than the true payment error rate**, a problematic situation even if occurring with very low probability. An extrapolated payment error rate that is larger than the true payment error rate, when extrapolated over an MAO's RADV-eligible population, would expose MAOs to significant financial risk based not on MAO coding accuracy but rather on the volatility of the CMS RADV payment error calculation methodology.

- CMS has yet to release information on the magnitude or derivation methodology of its **FFS Adjuster** offset to RADV payment penalties. The FFS Adjuster is intended to account for the fact that the documentation standard used to develop the MA risk adjustment model is inconsistent with the documentation standard used in RADV audits. Since CMS has yet to release details on its derivation, our technical analyses exclude consideration of the FFS Adjuster.

In summary, while our simulation work indicates that the CMS RADV sampling approach has the potential to accurately approximate the payment error rate of a contract, the payment error extrapolation approach exposes MA contracts to materially inequitable treatment based on characteristics independent of coding accuracy. The method is also exposed to the risk of unintended and problematic consequences such as payment penalties larger than actual payment error rates, albeit with low probability. Such sources of bias and inequity exist independent from

the to-be-defined FFS Adjuster. Although the FFS Adjuster would directionally mitigate the RADV financial risk exposure to MA contracts, as currently contemplated it would not lessen the bias and inequity evident in CMS's extrapolation approach. We detail our technical evaluation methodologies and findings in subsequent sections of this report.

Introduction

The Medicare Advantage (MA) program has long-relied on risk adjustment to ensure that payments to Medicare Advantage Organizations (MAO) reflect the relative health care risk of MAO beneficiaries. A well-designed risk adjustment system facilitates the alignment of plan payments with expected medical claims costs and therefore helps to support equity for Medicare beneficiaries in seeking coverage. The Centers for Medicare & Medicaid Services (CMS) relies on MAOs to regularly submit beneficiary diagnosis data to substantiate the health care risk of beneficiaries. CMS reserves the right to audit such substantiation to ensure payment accuracy. In 2012 CMS proposed a Risk Adjustment Data Validation (RADV) methodology that comprised a beneficiary sampling approach as well as a payment error calculation that involves extrapolation beyond the sampled audit population. The financial stakes of such an approach are significant since payment errors observed on a subset of MA beneficiaries could be used to levy financial penalties across a much larger population of MA beneficiaries. In this report we provide a brief background on MA risk adjustment, an overview of the CMS 2012 published RADV methodology, and an evaluation of the risks associated with the methodology. As detailed in this report, we found that the CMS sampling approach has the potential to approximate actual coding error rates, but the payment error extrapolation approach may expose MA contracts to materially inequitable treatment based on characteristics independent of coding accuracy.

America's Health Insurance Plans (AHIP) engaged Wakely Consulting Group (Wakely) to perform a technical evaluation of CMS's RADV payment error calculation methodology. This report summarizes the approach and key findings of the technical evaluation performed by Wakely.

Background on MA Risk Adjustment

MAOs receive monthly capitated payments from CMS that are adjusted to reflect the health care risk of enrolled beneficiaries. CMS uses a prospective risk adjustment system whereby MAOs submit health care diagnosis data to substantiate the risk profile of enrolled members. MAO-submitted diagnoses directly influence the risk scores assigned to MAO enrolled members, and in turn directly influence monthly CMS payments to MAOs. CMS calibrates the MA risk score model by correlating categories of diagnosis codes called Hierarchical Condition Categories (HCCs) with expected health care costs. Using a regression model that correlates demographic factors and HCCs to expected claims costs, each HCC is assigned a risk score value or "coefficient." If a beneficiary is identified via diagnosis code as being afflicted with a particular condition, the applicable HCC is triggered which may increase the risk score, and therefore the CMS payment, assigned to the beneficiary. Demographic characteristics, Medicaid eligibility status, and comorbidities among HCCs are among the numerous characteristics that the CMS risk adjustment model endeavors to account for in its payments to MAOs.

For perspective, a beneficiary with an average health care risk profile is assigned a risk score of 1.0, whereas a beneficiary with a risk score of 2.0 would be expected to have costs twice as high as the average beneficiary. An optimal risk adjustment program attempts to correlate beneficiary-specific funding with beneficiary-specific risk. This helps to ensure that financial resources are appropriately directed to plans that enroll complex, chronically ill, and costly beneficiaries.

Complete and accurate documentation of beneficiary diagnoses is an important component of the MA risk adjustment system. MAOs and CMS both invest resources to ensure that diagnosis codes submitted to CMS are accurate and appropriately documented, which supports the goals of payment accuracy and optimizing clinical care management.

RADV Overview

CMS conducts RADV audits on MA contracts as part of its program integrity efforts. Per CMS, MA RADV audits evaluate "whether the diagnosis codes submitted by MAOs can be validated by supporting medical record documentation."[1] In February of 2012 CMS published incomplete details of its RADV payment error calculation methodology, which comprises a member sampling method and a payment error extrapolation method.[1] CMS indicated that the extrapolation method would apply for the first time to RADV contract-level audits conducted on payment year 2011. Since publishing the methodology in early 2012, CMS has also conducted RADV audits on 2012 and 2013 payment years but has not yet released complete details on the payment error penalty calculation. Notably, the component of the methodology yet to-be-determined is a Fee-for-Service (FFS) Adjuster that is intended to account for the fact that the documentation standard used to develop the MA risk adjustment model is inconsistent with the documentation standard used in RADV audits. Refer to Appendix B for a more comprehensive summary of CMS's RADV methodology. Below we summarize the key elements.

1. CMS selects a set of approximately thirty (30) MA contracts for each RADV audit cycle (calendar year).

2. Within each selected contract members are flagged as "RADV-eligible" by satisfying a number of specific criteria, including the requirement that the member has at least one diagnosis code that resulted in the assignment of an HCC for the payment year.

---

[1] Notice of Final Payment Error Calculation Methodology for Part C Medicare Advantage Risk Adjustment Data Validation Contract-Level Audits. Available at: 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

3. CMS orders the RADV-eligible contract population by payment year risk score and divides the population into three equal-size strata (equal in terms of the number of beneficiaries in each stratum).

4. CMS randomly selects sixty-seven (67) enrollees from each of the three strata.

5. MAOs submit detailed medical records to support the HCCs represented in the RADV sample selected.

6. Based on the medical record documentation submitted, CMS calculates a RADV-corrected risk score and corrected payment amount.

7. CMS derives the MA contract-level payment error (penalty) by extrapolating the sample observed payment error to the entire RADV-eligible population – the extrapolation uses the lower bound of a ninety-nine (99) percent confidence interval around the estimated sample payment error.

8. The payment error is constrained to zero (0) if negative, meaning that CMS intends to recoup initial net overpayments, but does not intend to correct initial net underpayments.

9. CMS has indicated the intent to reduce any positive (non-zero) payment penalties by a to-be-defined FFS Adjuster.

Payment penalties are derived based on the lower bound of a ninety-nine (99) percent confidence interval around the observed sample payment error. While this approach materially reduces the derived payment penalty as compared to the observed sample payment error, it also introduces significant volatility into the calculation that, as we demonstrate via simulation, may potentially result in inequitable treatment of RADV audited contracts. Other aspects of the extrapolation methodology lead to results that are sensitive to sampling and subject to a high degree of randomness, which indicate additional, potentially problematic consequences of the CMS approach, as described further below.

Evaluation of CMS RADV Methodology

Goals

The principal goal of our technical evaluation of CMS's RADV methodology was to answer, through simulation, a few key questions:

1. Does the CMS sampling approach accurately estimate the simulated payment error rate of an MA contract?

2. Do contract attributes other than coding error rates influence the payment error calculation in a manner that drives potential inequitable treatment of contracts? Examples of such attributes:

   a. Contract size – number of enrollees

   b. Diagnostic mix (HCC profile) of MA contract population

   c. Diagnostic mix (HCC profile) of sample population

   d. Average risk score of the MA contract population

   e. Average risk score of the sample population

3. Is the CMS extrapolation approach exposed to the risk of unintended consequences, such as payment penalties that are larger than the actual payment error rate?

Evaluation Approach

In order to perform a data-driven evaluation of CMS's RADV sampling and extrapolation methodologies, we utilized the 2013 Medicare Limited Data Set (LDS) Standard Analytical Files and CMS HCC model coefficients[2] used to risk adjust 2013 MA payments. The LDS data comprises diagnostic information for approximately eight hundred thousand (800,000) Medicare beneficiaries that account for over two million HCCs. The publicly-available LDS data is distinctive in that it represents the HCC prevalence and mix of the Medicare-eligible population. We utilized the 2013 LDS data and MA payment year 2013 HCC coefficients to ensure alignment of payment year with risk adjustment model year.

We replicated the CMS RADV sampling and payment error extrapolation procedures on the LDS data using Monte Carlo simulation to first derive mock MA contracts of varying size (by enrollment), and then to mimic the CMS RADV methodology on such randomly generated contracts. Monte Carlo simulation, sometimes known as probability simulation, is a mathematical technique that uses random sampling to model the probability of different outcomes in a process. Monte Carlo simulation is a particularly valuable approach when applied to measuring uncertainty in processes that are impacted by random variables. This is an appropriate method for simulating the Medicare Advantage RADV process since there are several random variables involved, most notably: which two hundred and one RADV-eligible contract beneficiaries are randomly sampled, the diagnostic profile (HCCs) of the randomly selected beneficiaries in the sample, and the contract-specific coding error rates. We explored various scenarios of contract size and assumed coding error rates. For each combination of contract size and assumed coding error rate

---

[2] To simplify the analysis, disease interactions/comorbidities and disabled status/disease interactions were excluded from all risk scores and coding events. Additionally, lower-severity conditions that are "trumped" by more severe manifestations of the same disease hierarchy are excluded from the analysis. For example, a member with metastatic lung cancer will get scored for metastatic cancer (HCC8) and not lung cancer (HCC9) due to the cancer acuity hierarchy within CMS's model. If, upon RADV audit review, the metastatic cancer diagnosis is found to be unsubstantiated, it is possible that the member would be re-scored with the lower acuity lung cancer diagnosis (vs. removing the cancer diagnosis completely).

considered, we first randomly selected members to constitute simulated MA contract. We then simulated the RADV audit process one hundred thousand times (100,000) and derived summary statistics on the simulated RADV payment error penalties. Please refer to Appendix A for a more detailed description of the step-by-step Monte Carlo simulation approach utilized in our technical evaluation.

As CMS has yet to publish details on the FFS Adjuster (magnitude or derivation methodology), our scenarios do not assume a FFS Adjuster penalty offset. Note also that Wakely did not evaluate CMS's approach to selecting contracts for RADV audits – to our knowledge the methodology is not public. Finally, note that Wakely did not perform a comprehensive evaluation of CMS's operational approach to conducting RADV audits, including the required documentation and administrative procedures. Instead our evaluation focuses on the statistical elements of CMS's approach and the uncertainty inherent therein.

Note that CMS applies its payment error calculation to MAO contract-specific payments. Since our approach involved "simulated" MA contracts, we assume an average "standardized" (1.0 risk score) payment of $850 per member per month (PMPM) for purposes of dollarizing illustrative payment penalties throughout this report.

Key Findings

Our RADV simulation work yielded several key findings which are outlined below. A consistent theme in our findings is that CMS's RADV payment error extrapolation approach is prone to risk of inequitable treatment of contracts that vary in enrollment size, HCC mix, and absolute risk score. Refer to Appendix A for a broad range of scenarios tested and the statistical characteristics of resulting payment penalties.

Payment Error Penalties Are Subject to a High Degree of Randomness

CMS's sampling approach has the potential to accurately approximate the simulated payment error rate, and typically does so with a low variance. Despite the CMS approach's capability of approximating the payment error rate, the extrapolated **RADV payment penalties are subject to a high degree of randomness**. Small samples (201), combined with the fact that coding errors are somewhat rare contribute to erratic penalty results even at the same assumed true coding error rate. The observed randomness in payment penalties remains despite CMS's stratification approach, which contributes, albeit insufficiently, to more stable results as compared to an un-stratified sampling methodology.

For a tangible example, please see Table 1 below (duplicated in Appendix A). For three different contract sizes tested, we randomly sampled two hundred and one (201) beneficiaries one hundred thousand (100,000) times (three hundred thousand (300,000) scenarios tested in total for this "batch" of samples). For this set we assumed that ten (10) percent of HCCs were

unsupported (coded when they should not have been) and six (6) percent of HCCs were supported but not reported (not coded when they should have been). Note that the scenario parameters are loosely based on a Fiscal Year (FY) 2016 Department of Health and Human Services (HHS) Agency Financial Report.[3] This report reflects an average gross payment error rate of approximately ten (10) percent and an average net payment error rate of approximately four (4) percent that takes into account supported but not reported codes. The HHS report reflects data through June of 2015, and implies that supported but not reported coding errors represent a material offset to reported but unsupported coding errors. As an illustrative example, if we assume an average RADV sample beneficiary risk score of 1.04 and a standardized bid of $850, and we assume that the net payment error rate is four (4) percent, we would expect the 'true' risk score to be 1.0 and the average payment penalty PMPM to be:

($850 PMPM * 1.04) – ($850 PMPM * 1.0) = $34 PMPM

Table 1: Monte Carlo Simulation Assuming
10% HCCs Unsupported, 6% Supported But Not Reported

Contract RADV-Eligible Enrollment	PMPM Average Sample Payment Error	PMPM Average Sample Payment Error Variance	PMPM Payment Error Average Penalty	PMPM Min Penalty	PMPM Max Penalty	% Penalties > 0	% High Penalties
1,000	$33.55	$0.26	$2.26	$0.00	$46.80	26.8%	0.102%
10,000	$33.58	$0.33	$2.60	$0.00	$67.07	27.6%	0.238%
100,000	$32.43	$0.33	$2.66	$0.00	$57.98	27.6%	0.326%

---

[3] Department of Health and Human Services, Fiscal Year (FY) 2016 Department of Health and Human Services (HHS) Agency Financial Report. Available online at: https://www.hhs.gov/sites/default/files/fy-2016-hhs-agency-financial-report.pdf

For reference we define the column headings below from Table 1 below (repeated in Appendix A):

Table 1 Column Heading	Definition
Contract RADV-eligible Enrollment	Number of beneficiaries in the simulated MA contract – each row in the tables represents a unique set of scenarios
PMPM Average Sample Payment Error	The per member per month (PMPM) average value of the simulated sample coding errors
PMPM Average Sample Payment Error Variance	The PMPM value of the variance of sample payment errors
PMPM Payment Error Average Penalty	The average PMPM value of the extrapolated payment penalties
PMPM Min Penalty	The minimum extrapolated penalty among all RADV simulations within the scenario set
PMPM Max Penalty	The maximum extrapolated penalty among all RADV simulations within the scenario set
% Penalties > 0	The percentage of RADV scenarios in which the payment penalty is greater than $0
% High Penalties	The percentage of RADV non-zero penalty scenarios in which the extrapolated penalty is larger than the true payment error rate

For each of the simulated contract sizes, we pulled one hundred thousand (100,000) random RADV samples, replicated the RADV payment penalty calculation, and derived a number of summary statistics. For a particular contract size, we derived the sample average payment error, the variance in payment penalty, the minimum penalty, and maximum penalty, the percentage of scenarios for which a nonzero penalty was generated, and the percentage of scenarios for which the penalty was larger than the assumed true payment error rate (in this case approximately 4 percent net error rate). As detailed in Appendix A, we modeled a wide array of true error rate simulations, but our key findings generally hold true across all scenarios tested.

The sampling approach has the potential to accurately approximate the value of the expected payment error – based on a ten (10) percent unsupported coding error rate and a six (6) percent supported but not reported coding error rate, we would expect approximately a $34 PMPM error (0.04 risk score value), with very low variance. However, the payment penalties do yield a wide range of outcomes when simulating the RADV process on the same contract – ranging from no penalty ($0) to penalties that are approximately double ($67.07 PMPM) the average sample payment error ($33.58 PMPM). Such variation in payment penalty for randomly chosen RADV samples drawn from the same contract is obviously problematic. Our simulation exercise illustrates that if CMS runs the RADV process on the same contract twice, the resulting payment penalties may vary significantly. The instability in simulated payment penalty results suggests that the RADV extrapolation process cannot reliably and equitably align payment penalties with actual payment error rates.

Payment Penalties May be Higher than the True Coding Error Rate

We emphasize that **payment penalties derived by CMS's methodology have the potential to be materially higher than the true payment error rate,** albeit with low probability. For example, refer to Table 1 above - the maximum penalty observed in the ten thousand (10,000) contract enrollment scenarios was $67.07 PMPM, nearly double the average risk score error observed over one hundred thousand (100,000) samples of $33.58 PMPM. Note that we define the "High penalty" in Table 1 as a penalty that is higher than the modeled true error rate of the contract. The nonzero probability of such a penalty that is higher than the true payment error is problematic, and could lead to a situation where RADV-audited contracts forfeit material funding due to randomness in CMS's sampling methodology, not due to coding accuracy.

CMS Methodology May Inequitably Penalize Contracts as Enrollment Increases

As previously noted, we define "High penalty" cases as those for which the CMS-derived penalty is greater than the true error rate. As can be seen in Table 1 and across the multitude of scenario "sets" tested in Appendix A, **larger contract sizes are generally penalized by greater randomness in penalties**. This is particularly evident when looking at the percentage of scenarios that yield a "High penalty" – a metric that generally increases with contract size based on our simulation work.

Notice in Table 1 that both the average PMPM payment error penalty and the percentage of "High penalties" increases with contract size by enrollment. RADV sample size does not vary by contract size (for contracts with at least one thousand RADV-eligible beneficiaries). If two contracts of very different sizes have identically-distributed errors over the entire population, there will tend to be more variance (i.e. more randomness) in the penalties drawn from the larger contract due to the fixed sample size.

Payment Error Penalties Are Sensitive to Small Variations in Sample Population HCCs

Our simulation work affirmed what we believe to be an intuitive observation: **the HCC profile of the randomly selected RADV sample may drive significant variation in the payment penalty**. In other words, the diagnoses/HCCs that the randomly selected RADV sample members happen to have may drive material variation in payment penalty. The CMS methodology randomly selects RADV-eligible beneficiaries, and each beneficiary may have a significantly different mix of HCCs. Below see an example of how the inclusion or exclusion of a single HCC error from the sampled RADV population drives a significant change in payment penalty. Using one of the actual RADV samples simulated in our Monte Carlo work, we measured the sensitivity of the RADV payment error penalty to the inclusion of a single additional HCC coding error.

Figure 1: Sensitivity of RADV Penalty to Single Coding Errors

RADV Scenario Parameters	
RADV Sample Members	201
Contract Enrollment Size (RADV-eligible)	100,000
Simulated probability of unsupported coding error	10%
Simulated probability of supported but not reported coding error	6%
Simulated net coding error rate	4%
Assumed MA Standardized Bid (1.0 Risk Score) PMPM	$850.00
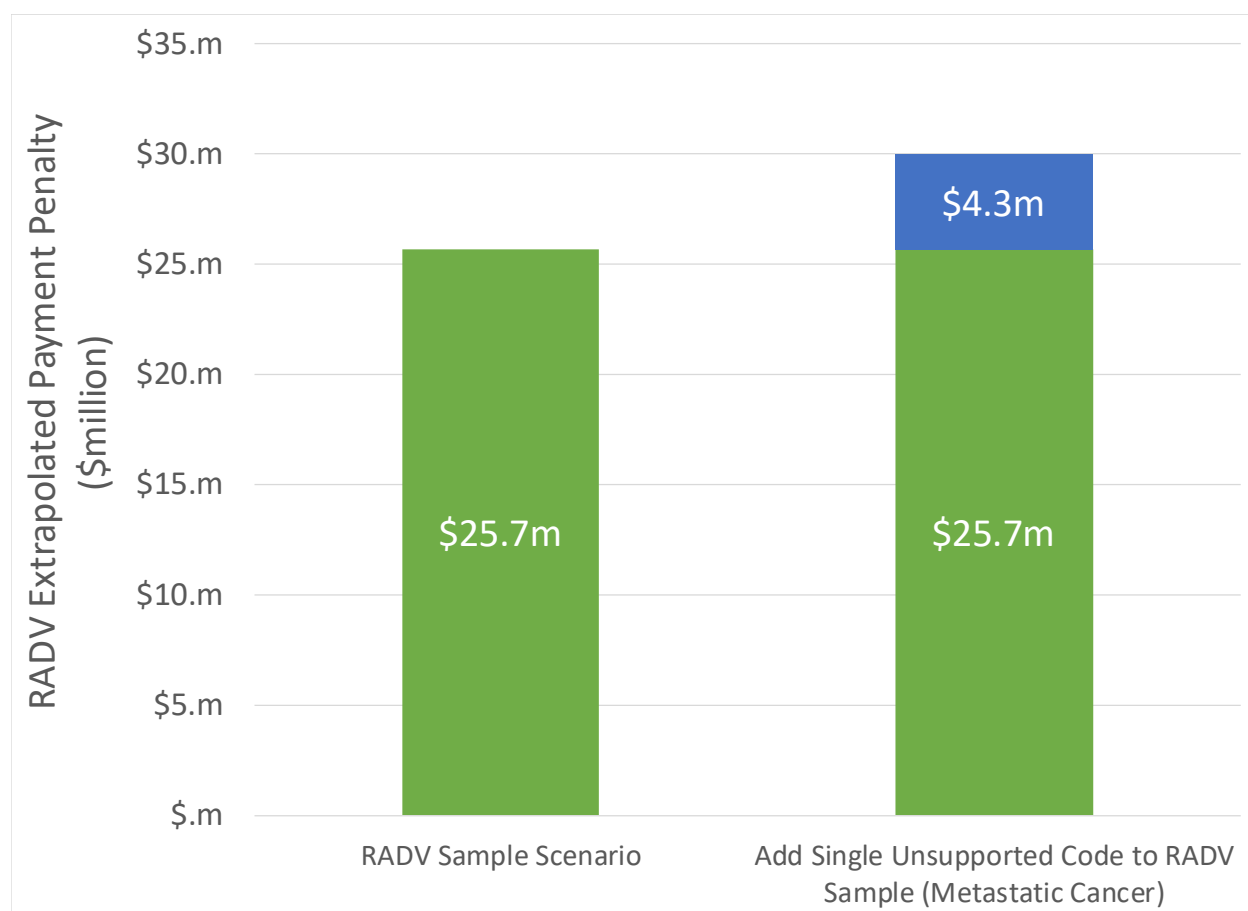
As illustrated above, the sensitivity of CMS's payment penalty calculation to a single HCC error is problematic in that random chance could drive material swings in extrapolated payment penalties. For this particular example, adding a single unsupported coding error (metastatic cancer) to the RADV sample results in a 16.7 percent ($4.3 million) increase in the RADV penalty. Random chance associated with a single high acuity condition being present in a RADV audit

sample exposes MA contracts to financial penalties that vary not based just on coding accuracy, but also on the "luck of the draw" in the randomly selected sample population.

CMS Methodology May Treat Two Contracts with Identical Payment Error Rates Differently

The CMS approach of deriving the RADV payment penalty from the lower bound of a confidence interval **may drive inequitable treatment of contracts with identical average error rates**. As validated by our scenario testing, two contracts with virtually identical payment error rates may be subject to vastly different penalties. More specifically, for two issuers with the same payment error rate but with different observed error variances, the issuer with the greater observed error variance (i.e. more volatility) will face the lower penalty.

Below we illustrate an example of two simulated RADV scenarios (actual scenarios from our simulation work) that reflect the same simulated true payment error rate ($32.41 PMPM), the same observed sample average error rate ($75.32 PMPM), but materially different RADV penalties. The disparity in penalties is driven by the difference in RADV sample error variance between the two scenarios. Note that these examples also illustrate that the random chance element of the RADV sampling process may yield a sample payment error ($75.32) materially larger than the true payment error ($32.41). For the specific examples summarized in Figure 2, scenario 99914 reflects a standard error (square root of variance) PMPM value of $24.46, whereas the standard error for scenario 22046 is $13.36 PMPM. This means that the RADV sample drawn for scenario 99914 reflects more volatility, or a larger variation from the sample average error rate, as compared to the sample drawn for scenario 22046. Since the CMS RADV payment penalty reflects the lower bound of the ninety-nine (99) percent confidence interval around the observed sample error, higher variation (higher standard error) drives a lower payment penalty.

Figure 2: Variation in RADV Penalties for Contract with Identical Average Errors

RADV Scenario Parameters	
RADV Sample Members	201
Contract Enrollment Size (RADV-eligible)	100,000
Simulated probability of unsupported coding error	10%
Simulated probability of supported but not reported coding error	6%
Simulated net coding error rate	4%
Assumed MA Standardized Bid (1.0 Risk Score) PMPM	$850.00
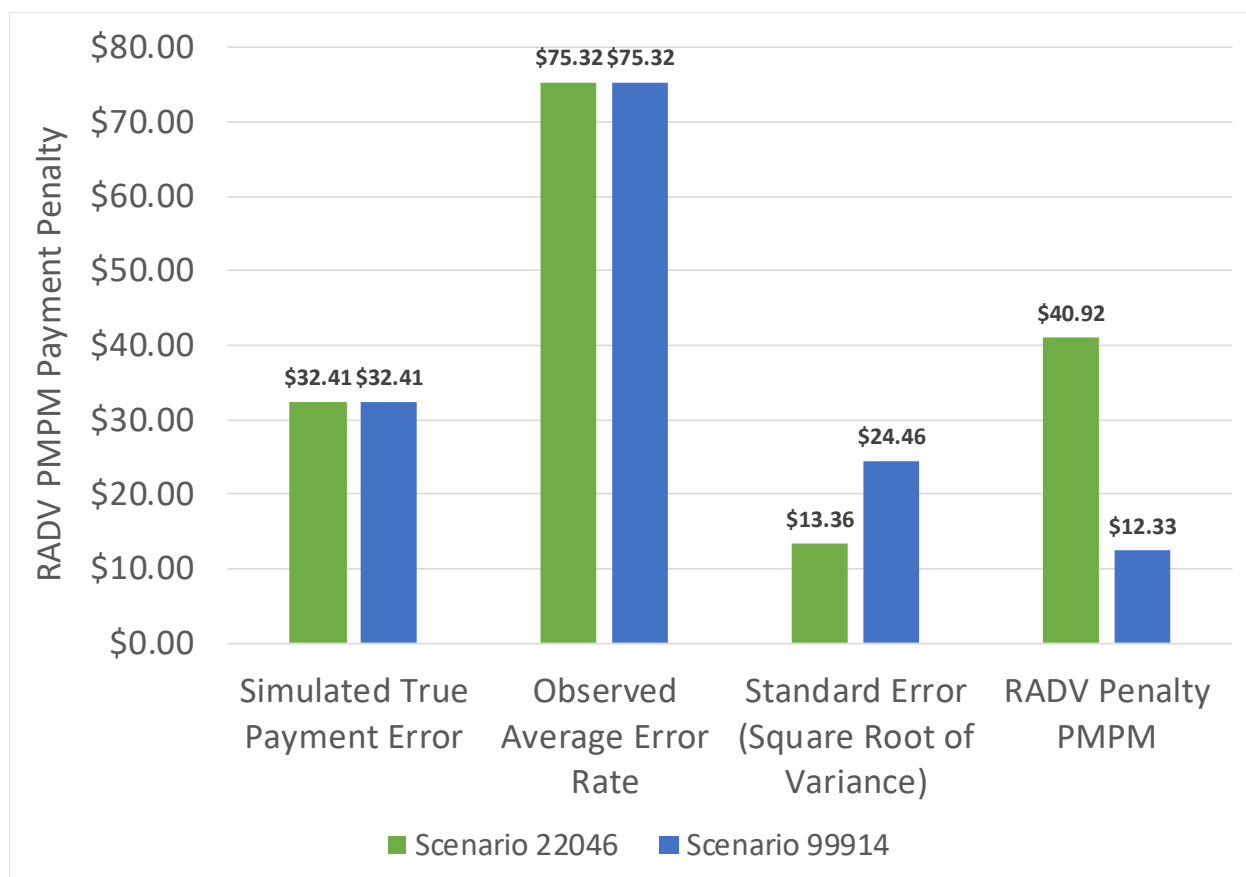
CMS Methodology May Inequitably Penalize Low Risk Score Contracts

We note from our simulation work that the CMS RADV methodology may drive **disproportionate and potentially inaccurate penalization of low risk score contracts**. If two issuers have the same true payment error per 1.0 risk score value[4] but materially different absolute average risk scores, then the variance of the errors in the contract with the higher risk score will likely have a lower penalty per 1.0 risk score value than the contract with the lower risk score. This inequitable treatment of contracts by absolute risk score is related to the confidence interval approach that CMS uses to calculate payment penalties. Since CMS uses the lower bound of the confidence interval to define the payment penalty, greater variance in the observed payment errors drives a lower penalty result.

---

[4] For example, a five (5) percent error rate per 1.0 of risk score would mean that a contract with an average risk score of 2.0 would have an expected risk score value error of 0.1.

Other Comments on CMS Methodology

The CMS Methodology does not consider HCC-specific substantiation rates. CMS's RADV process randomly selects beneficiaries and each selected beneficiary is assigned the same weight in the extrapolation penalty calculation. However, each sampled beneficiary has a unique HCC profile, or "mix" of HCCs. Further, each HCC has its own substantiation success rate within the industry, as some HCCs are harder to substantiate than others. Therefore, depending on the HCC profile of the contract population, as well as the RADV sample population, there may be material variation in "expected" coding error rates.

CMS acknowledges its understanding that error rates may vary by HCC, as illustrated in several recent publications:

- Proposed [2019 Notice of Benefit and Payment Parameters](#)[5] (page 51074) - "*HHS could also evaluate error rates within each HCC, or groups of HCCs, then only apply error rates to outlier' issuers' risk scores within each HCC or group of HCCs.*"

- Final [2019 Notice of Benefit and Payment Parameters](#)[6] (page 16961) – "*Our simulations of failure rates by HCC group suggest that such an approach yields a more equitable measure to evaluate statistically different HCC failure rates affecting an issuer's error rate than an approach based on an overall failure rate, which may overly adjust issuers with abnormal distributions of certain HCCs due to their underlying populations rather than differences due to errors in diagnoses codes.*"

- [December 2015 Statement of Work for RADV Recovery Audit Contractors](#)[7] - "*Condition Specific RADV Audits will be conducted for a subset of MA contracts not subject to a Comprehensive Audit for any given payment year. The focus of Condition Specific Audits will be a set(s) of HCCs determined to have a higher probability of being erroneous, for example, it may be decided that the hierarchy of HCCs relating to 'diabetes' should be the subject of this targeted review.*"

The MA RADV audit standard against which coding error rates are measured is not reflective of varying expected error rates by HCC. This is problematic for a few reasons. First, not considering HCC-specific substantiation rates virtually guarantees inequitable treatment of two different

---

[5] Patient Protection and Affordable Care Act; HHS Proposed Notice of Benefit and Payment Parameters for 2019. Available online at: https://www.gpo.gov/fdsys/pkg/FR-2017-11-02/pdf/2017-23599.pdf

[6] Patient Protection and Affordable Care Act; HHS Final Notice of Benefit and Payment Parameters for 2019. Available online at: https://www.gpo.gov/fdsys/pkg/FR-2018-04-17/pdf/2018-07355.pdf

[7] The Medicare Advantage (MA) Risk Adjustment Data Validation (RADV) Recovery Audit Contractor (RAC) Request for Information. Available online at: https://www.fbo.gov/utils/view?id=e50f5bb5f02c9fc7d9815f163f0941a4.

contracts selected for RADV audit. Both contracts will be held to the same coding standard despite the fact that varying HCC profiles would drive varying coding error rate expectations. Second, even within the same contract, randomly selected samples of beneficiaries will yield varying HCC profiles, which should be measured against HCC-specific substantiation standards. Therefore, two different samples from the same contract could yield materially different payment penalties as a result of varying HCC profiles of two randomly selected samples. As demonstrated in Figure 1, even a single coding error from a RADV sample can drive material variance in payment penalties. Therefore, widely varying HCC profiles of two randomly selected samples may drive material variance in payment penalties based not on coding accuracy but rather on HCC mix.

CMS has not yet released details on its FFS Adjuster. As noted by CMS in its February 2012 methodology release, CMS calibrates the MA risk score model using Medicare FFS claims experience. CMS acknowledges that the coding documentation standard used in RADV audits is different from the coding documentation standard used to develop the MA risk adjustment model. The FFS adjuster is intended to account for this disconnect, since it would be mathematically inconsistent to hold MAOs to a stricter coding standard than that used to develop the MA risk adjustment model.[8] The details of the FFS Adjuster, its magnitude and derivation methodology, have not yet been released. Therefore, an evaluation of the FFS Adjuster is not possible at this point.

CMS excludes members with zero HCCs ("zero HCC" or "non-HCC") from the RADV-eligible contract population, which biases the sample payment error rate upwards. Any supported but not reported codes on the population that is not RADV-eligible are systematically ignored in the CMS approach. A Fiscal Year (FY) 2016 Department of Health and Human Services (HHS) Agency Financial Report reflects an average gross payment error rate of approximately ten (10) percent and an average net payment error rate of approximately four (4) percent that takes into account supported but not reported codes. The HHS report reflects data through June of 2015, and implies that supported but not reported coding errors represent a material offset to unsupported coding errors.

Excluding non-HCC members from the RADV audit samples biases the observed payment error by removing potential supported but not reported codes for non-HCC members. This makes the expected observed payment error rate higher than the true payment error rate over the entire contract (RADV-eligible plus non-eligible). Supported but not reported codes for non-HCC beneficiaries are completely unaddressed in the CMS RADV methodology since these beneficiaries are excluded from the population from which members are sampled.

---

[8] American Academy of Actuaries. "Re: Comment on RADV Sampling and Error Calculation Methodology." Received by Cheri Rice, 21 January 2011. Available online at: https://www.actuary.org/pdf/health/RADV_comment_letter_012111_final.pdf

CMS operational processes may exclude supported but not reported coding errors from the calculation of the payment error. While we do not endeavor to provide a comprehensive review of RADV audit operational design, we do note that key operational parameters could serve to influence penalty calculation outcomes. One of these operational parameters involves medical chart submissions. For each HCC that CMS attempts to validate via RADV audit, MAOs are permitted to submit up to five charts that (potentially) substantiate the HCC in question. The RADV auditor identifies the first chart that substantiates the HCC, and then codes every diagnosis on that particular chart. The first substantiating chart may uncover supported but not reported codes which would be accounted for in CMS's approach, but supported but not reported codes from the remaining charts would not be uncovered. If none of the charts substantiate the HCC, then it is not clear that un-submitted diagnoses from any of the five charts would be coded, which would eliminate any chance to uncover supported but not reported conditions. Instead of ensuring a comprehensive and accurate picture of a selected beneficiary's health care status, the operational limitations of the RADV audit process potentially restrict CMS from gaining a complete picture of coding errors (both unsupported codes and supported but not reported codes). Therefore, the potential exclusion of some medical charts from the RADV audit process biases the observed payment error upwards when compared to the true payment error. Such potential overstatement of the true payment error could inflate the observed sample payment error and therefore erroneously inflate extrapolated payment penalties.

Conclusion

CMS's payment error extrapolation approach exposes MA contracts to materially inequitable treatment based on characteristics independent of coding accuracy. While the random sampling approach has the potential to accurately approximate payment error rates, the design of the payment error extrapolation calculation introduces the risk of unintended and problematic consequences such as larger payment penalties for contracts with low variance in coding errors. The inherent randomness in the HCC profile of a RADV sample, as well as CMS-acknowledged variations in HCC-specific substantiation rates, further contribute to the potential for inequitable treatment by contract. Such sources of bias and inequity exist independently from the to-be-defined FFS Adjuster. Although the FFS Adjuster would mitigate the absolute value of financial risk exposure to MA contracts, there is currently no evidence to suggest that it would lessen the bias and potential inequity evident in CMS's extrapolation approach.

Considerations and Limitations

Wakely was commissioned by America's Health Insurance Plans (AHIP) to perform a technical evaluation of the CMS RADV methodology. The report should be considered in its entirety. The report represents a technical evaluation and does not represent support for any particular policy or changes thereof. We do not intend this information to benefit any third party nor create a reliance by any third party on Wakely. Wakely is not responsible for any use of the report or

consequences of such use outside the specific purpose for which it was intended. Any mathematical estimates included in this report and produced by our Monte Carlo simulation exercise are inherently uncertain. Users of the report results should be qualified to use it and understand the results and the inherent uncertainty.

Appendix A – Monte Carlo Simulation Background and Results

Monte Carlo Simulation Background

Monte Carlo simulation, sometimes known as probability simulation, is a mathematical technique that uses random sampling to model the probability of different outcomes in a process. Monte Carlo simulation is a particularly valuable approach when applied to measuring uncertainty in processes that are impacted by random variables.

Monte Carlo simulation is an appropriate method for simulating the MA RADV process since there are several random variables involved, most notably: which two hundred and one randomly selected RADV-eligible contract beneficiaries are sampled, the diagnostic profile (HCCs) of the randomly selected beneficiaries in the sample, the contract-specific true unsupported coding error rate, and the contract-specific true supported but not reported coding error rate. By holding constant the contract size and assumed true coding error rates for particular "scenario sets[9]," one can evaluate over a large number of samples the statistical attributes of CMS's sampling and extrapolation methodology. More specifically, one can evaluate how closely the observed coding errors track the true coding error rate, the probability of nonzero payment penalties, whether there are risks of biases in the extrapolation calculations, the frequency and severity of unintended extrapolation results, and other statistical attributes. While it is not possible to mimic all operational aspects of CMS's RADV audit approach on actual Medicare Advantage coding data, the deployment of Monte Carlo simulation on actual Medicare beneficiary diagnostic data enables a robust mathematical evaluation.

Monte Carlo Simulation Approach

We started by limiting the 2013 Medicare LDS data set to beneficiaries that satisfy criteria for RADV eligibility. We then simulated MA contracts of varying sizes by randomly selecting enrollees to make up three contract sizes – one thousand (1,000) enrollees, ten thousand (10,000) enrollees, and one hundred thousand (100,000) enrollees. For each contract size tested, we defined coding error rates (unsupported codes and supported but not reported codes) and randomly assigned actual coding errors to the RADV-eligible population diagnostic data. Note that we did not assume varying coding error rates by HCC.

---

[9] We use the term "scenario set" to refer to a particular combination of MA contract size and coding error rates assumed (e.g. one thousand (1,000) beneficiaries, ten (10) percent unsupported coding error rate, six (6) percent supported but not reported error rate).

- **Assumed unsupported coding error rate** – the probability that a particular HCC is erroneously reported but unsupported, i.e. submitted by the MAO as a valid diagnosis code despite the supporting documentation being insufficient

- **Assumed supported but not reported coding error rate** – the probability that a particular HCC is supported but erroneously not reported, i.e. not submitted by the MAO as a valid diagnosis code despite the supporting documentation being sufficient

We then replicated the RADV sampling process on the simulated MA contracts: we stratified the RADV-eligible contract population into three groups based on risk score, randomly selected sixty-seven (67) beneficiaries from each cohort, calculated the sample average errors and sample error variances, and finally derived an extrapolated payment penalty per the CMS-published formula. Note that a more complete summary of CMS's methodology is captured in Appendix B.

For each of the scenario sets explored, we simulated the RADV sampling and payment penalty extrapolation one hundred thousand (100,000) times. For each single scenario, we replicated CMS calculations to derive an extrapolated payment penalty (excluding the yet-to-be-defined Medicare FFS adjuster). For each scenario set we calculated the sample average risk score coding error, the variance in average errors of the samples, the average penalty as a percent of premium, the maximum/minimum penalties among scenarios tested, the percentage of scenarios that generated a nonzero positive penalty, and the percentage of penalties that were higher than the true error rate of the underlying contract (referred to as a "High penalty").

Refer to the tables below for summary statistics on a number of Monte Carlo simulations of RADV sampling and payment penalty calculations across varying contract sizes and assumed true error rates. We first define the table column headings that are used consistently across all six Appendix A scenario set summary tables:

Appendix A Table Column Heading	Definition
Contract RADV-eligible Enrollment	Number of beneficiaries in the simulated MA contract – each row in the tables represents a unique set of scenarios
PMPM Average Sample Payment Error	The per member per month (PMPM) average value of the simulated sample coding errors
PMPM Average Sample Payment Error Variance	The PMPM value of the variance of sample payment errors
PMPM Payment Error Average Penalty	The average PMPM value of the extrapolated payment penalties
PMPM Min Penalty	The minimum extrapolated penalty among all RADV simulations within the scenario set
PMPM Max Penalty	The maximum extrapolated penalty among all RADV simulations within the scenario set
% Penalties > 0	The percentage of RADV scenarios in which the payment penalty is greater than $0
% High Penalties	The percentage of RADV non-zero penalties in which the extrapolated penalty is larger than the true coding error rate

Table A1: 10% HCCs Unsupported, 6% Supported but Not Reported

Contract RADV-Eligible Enrollment	PMPM Average Sample Payment Error	PMPM Average Sample Payment Error Variance	PMPM Payment Error Average Penalty	PMPM Min Penalty	PMPM Max Penalty	% Penalties > 0	% High Penalties
1,000	$33.55	$0.26	$2.26	$0.00	$46.80	26.8%	0.102%
10,000	$33.58	$0.33	$2.60	$0.00	$67.07	27.6%	0.238%
100,000	$32.43	$0.33	$2.66	$0.00	$57.98	27.6%	0.326%

Table A2: 5% HCCs Unsupported, 5% Supported but Not Reported

Contract RADV-Eligible Enrollment	PMPM Average Sample Payment Error	PMPM Average Sample Payment Error Variance	PMPM Payment Error Average Penalty	PMPM Min Penalty	PMPM Max Penalty	% Penalties > 0	% High Penalties
1,000	-$4.08	$0.16	$0.00	$0.00	$9.75	0.1%	100.000%
10,000	-$0.49	$0.20	$0.01	$0.00	$14.17	0.4%	100.000%
100,000	-$0.13	$0.21	$0.02	$0.00	$19.67	0.4%	100.000%

Table A3: 30% HCCs Unsupported, 0% Supported but Not Reported

Contract RADV-Eligible Enrollment	PMPM Average Sample Payment Error	PMPM Average Sample Payment Error Variance	PMPM Payment Error Average Penalty	PMPM Min Penalty	PMPM Max Penalty	% Penalties > 0	% High Penalties
1,000	$242.61	$0.52	$182.59	$110.84	$262.94	100.0%	0.044%
10,000	$240.84	$0.63	$181.53	$107.66	$278.98	100.0%	0.133%
100,000	$243.48	$0.63	$184.59	$111.35	$277.17	100.0%	0.171%

Table A4: 10% HCCs Unsupported, 0% Supported but Not Reported

Contract RADV-Eligible Enrollment	PMPM Average Sample Payment Error	PMPM Average Sample Payment Error Variance	PMPM Payment Error Average Penalty	PMPM Min Penalty	PMPM Max Penalty	% Penalties > 0	% High Penalties
1,000	$80.19	$0.17	$46.53	$13.77	$86.52	100.0%	0.008%
10,000	$81.11	$0.21	$46.85	$12.21	$92.64	100.0%	0.052%
100,000	$81.00	$0.21	$46.94	$12.23	$94.05	100.0%	0.075%

Table A5: 5% HCCs Unsupported, 0% Supported but Not Reported

Contract RADV-Eligible Enrollment	PMPM Average Sample Payment Error	PMPM Average Sample Payment Error Variance	PMPM Payment Error Average Penalty	PMPM Min Penalty	PMPM Max Penalty	% Penalties > 0	% High Penalties
1,000	$38.32	$0.07	$16.43	$0.00	$42.36	100.0%	0.006%
10,000	$41.55	$0.12	$16.33	$0.00	$48.44	99.9%	0.035%
100,000	$40.64	$0.11	$16.79	$0.00	$48.23	99.9%	0.045%

Table A6: 1% HCCs Unsupported, 0% Supported but Not Reported

Contract RADV-Eligible Enrollment	PMPM Average Sample Payment Error	PMPM Average Sample Payment Error Variance	PMPM Payment Error Average Penalty	PMPM Min Penalty	PMPM Max Penalty	% Penalties > 0	% High Penalties
1,000	$7.45	$0.01	$0.19	$0.00	$7.71	20.4%	0.001%
10,000	$7.88	$0.02	$0.14	$0.00	$8.86	13.7%	0.001%
100,000	$8.02	$0.02	$0.15	$0.00	$9.64	14.2%	0.006%

Appendix B – CMS RADV Methodology

Sampling and Extrapolation Methodology[10]

In this section we paraphrase and summarize CMS's February 2012 Notice of Final Payment Error Calculation methodology for Part C MA RADV Contract-Level Audits.

ELIGIBILITY FOR RADV SAMPLING

CMS selects[11] a set of MA contracts for each RADV audit cycle. Within each contract selected for RADV audit, a sample of enrollees is defined selected in order for CMS to estimate a contract-level risk adjustment payment error. The sample enrollees are drawn from contract population that CMS deems "RADV-eligible" by virtue of meeting all of the following criteria:

1. Beneficiary is enrolled in the selected MA contract as of January of the diagnosis collection period (calendar year prior to payment year) and continuously enrolled for twelve (12) months through January of the payment year

2. Beneficiary is not identified by CMS as End-Stage Renal Disease (ESRD) status and is not identified as hospice status at any time from January of the diagnosis collection period through January of the payment year.

3. Beneficiary is enrolled in Medicare Part B coverage for all twelve (12) months of the diagnosis collection period.

4. Beneficiary has at least one diagnosis code submitted that led to the assignment of at least one HCC for the payment year.

SAMPLE SIZE AND STRATA

CMS orders the RADV-eligible contract population based on payment year risk score (lowest to highest) and divides the sample into three equal size groups, or strata. Sixty-seven (67) enrollees from each of the three strata are randomly selected, generating a total sample size of two hundred and one (201) enrollees. Note that smaller samples are drawn if a contract's RADV-eligible

---

[10] See Notice of Final Payment Error Calculation Methodology for Part C Medicare Advantage Risk Adjustment Data Validation Contract-Level Audits. Available online at: 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

[11] Thirty (30) contracts were selected for 2011 payment year MA RADV audits. This report does not include an assessment of CMS's methodology for selecting MA contracts for RADV audits, as the methodology has not been published to our knowledge.

population is less than one thousand (1,000) enrollees. At this stage, CMS defines a "**stratum-based enrollee weight**" as the number of enrollees in the stratum divided by the number of enrollees randomly selected (usually sixty-seven).

For example, if a contract has fifteen thousand (15,000) RADV-eligible enrollees, sixty-seven enrollees (67) would be randomly selected from each of three strata of five thousand (5,000) enrollees. The stratum-based enrollee weight in this case would be five thousand (5,000) divided by sixty-seven (67), or 74.627. The stratum-based enrollee weight is used as a multiplier to extrapolate the payment error measured to the entire RADV-eligible population of the stratum.

DOCUMENTATION SUBMISSION PARAMETERS

MAOs are required to submit detailed medical records to support all HCCs represented in the randomly selected beneficiary sample. CMS permits audited MA contracts to submit multiple medical records for each HCC being validated. However, all diagnoses identified in the first medical record that validates the HCC will be used. The "one best medical record" policy applies to the RADV audit dispute and appeal processes. In the event of a RADV audit dispute/appeal CMS requires that MAOs submit a single medical record that best substantiates the HCC in question.

PAYMENT ERROR CALCULATION

Based on the medical record documentation submitted, CMS calculates a RADV-corrected risk score and corrected payment amount. The risk score value of HCCs not substantiated by medical record documentation is removed from enrollee risk scores, and the HCC value of any previously undocumented diagnoses is added to the enrollee risk scores. Per member per month (PMPM) payment errors are defined as the difference between the original monthly CMS payment to the MAO and the RADV-corrected monthly payment for each enrollee. Note that payment errors at the enrollee level may be positive (overpayment to MAO) or negative (underpayment to MAO). CMS derives an annual payment error for each sampled enrollee by multiplying the PMPM payment error by the number of months the beneficiary was enrolled during the payment year.

PAYMENT ERROR EXTRAPOLATION

CMS derives the MA contract-level payment error by extrapolating the observed annual payment error to the entire RADV-eligible population. Put simply, CMS estimates the average payment error based on the randomly selected sample of beneficiaries and calculates a nine-nine (99) percent confidence interval (CI) around that estimated error. In other words, CMS is implying that there is a ninety-nine (99) percent chance that the actual payment error will be between the lower and upper bounds of its confidence interval.

The more intricate details of the extrapolation are outlined in the paragraph below:

CMS multiplies the annual payment error for each sampled enrollee by the stratum-based enrollee weight (previously defined). The extrapolated enrollee annual payment error is summed across all enrollees in the sample to derive an estimated "point estimate" of the contract-level payment error. Importantly, a ninety-nine (99) percent CI for the payment error is calculated for each RADV audited contract. The ninety-nine (99) percent CI is derived by varying the estimated payment error observed (average observed payment error) by 2.575 times the Standard Error (SE). The SE is derived as follows:

1. Derive the variance (v_h) of the unweighted enrollee payment errors within each of three strata (h).

2. Calculate the variance of the estimated total ($V_{T^\wedge}$) payment error where N represents the number of RADV-eligible enrollees in stratum h.

$$V_{T^\wedge} = \sum_{h=1}^{3} \frac{N_h^2}{67} \, v_h$$

3. $Standard\ Error\ SE_T = \sqrt{V_{T^\wedge}}$

PAYMENT RECOVERY AMOUNT AND FFS ADJUSTER

CMS uses the lower bound of the derived confidence interval to determine the payment recovery amount (the amount that CMS intends to recoup from the MAO). Note that we use "payment recovery amount" and "payment penalty" interchangeably throughout the report. CMS sets the recovery amount floor at zero. In other words, if the lower bound of the confidence interval below zero, which indicates that CMS may have underpaid the MAO initially, CMS will not make an incremental payment to the MAO to correct for the initial underpayment. If the lower bound of the derived payment error confidence interval is above zero, then the lower bound of the confidence interval will define the "preliminary payment recovery amount." This preliminary amount will be adjusted downward by a to-be-defined Fee-for-Service Adjuster (FFS Adjuster), but still constrained to the zero recovery floor.

The concept of the FFS Adjuster is intended to account for the difference in coding documentation standards between the MAO medical records and the FFS claim medical records used to develop the Medicare Advantage risk adjustment model. CMS has indicated that the FFS Adjuster will be derived based on a "RADV-like review of records submitted to support FFS claims data." To our knowledge, since the February 2012 release of the MA RADV Sampling and Extrapolation Methodology, CMS has yet to release any substantive information on the FFS Adjuster amount or its derivation.