JOSEPH H. HUNT
Assistant Attorney General, Civil Division
NICOLA T. HANNA
United States Attorney
DAVID M. HARRIS
DAVID K. BARRETT
ABRAHAM C. MELTZER
JOHN E. LEE (CBN 128696)
Assistant United States Attorneys
300 N. Los Angeles Street, Room 7516
Los Angeles, California 90012
Tel: (213) 894-3995; Fax: (213) 894-7819
Email: john.lee2@usdoj.gov
MICHAEL D. GRANSTON
DANIEL R. ANDERSON
EDWARD C. CROOKE
CAROL L. WALLACK
JUSTIN DRAYCOTT
PAUL G. FREEBORNE
JESSICA KRIEG
ZOILA E. HINSON
AMY L. LIKOFF
ADAM R. TAROSKY
Attorneys, Civil Division, U.S. Department of Justice
P.O. Box 261, Ben Franklin Station
Washington, D.C. 20044
Tel: (202) 307-0486; Fax: (202) 307-3852
Email: carol.wallack@usdoj.gov
JAMES P. KENNEDY, JR.
United States Attorney
KATHLEEN ANN LYNCH
Assistant United States Attorney (Admitted PHV)
138 Delaware Avenue
Buffalo, New York 14201
Tel: (716) 843-5830; Fax: (716) 551-3052
Email: kathleen.lynch@usdoj.gov
Attorneys for the United States of America

UNITED STATES DISTRICT COURT

FOR THE CENTRAL DISTRICT OF CALIFORNIA

WESTERN DIVISION

UNITED STATES OF AMERICA *ex rel*. BENJAMIN POEHLING, Plaintiffs, v. UNITEDHEALTH GROUP, INC., *et al*., Defendants.	No. CV 16-08697 MWF (SSx) NOTICE OF RECENT PROPOSED MEDICARE PART C REGULATION RELATING TO FEE-FOR-SERVICE ADJUSTER ISSUE; EXHIBITS Date: September 17, 2018 Time: 10:00 a.m. Court: 5A, Hon. Michael W. Fitzgerald

PLEASE TAKE NOTICE that, on October 26, 2018, the Department of Health and Human Services issued proposed revisions to its Medicare Part C Risk Adjustment Data Validation (RADV) Audit regulations along with comments addressing the Fee-For-Service (FFS) Adjuster issue and also posted the following related documents on the Centers for Medicare and Medicaid Services (CMS) website: an Executive Summary entitled Fee for Service Adjuster and Payment Recovery for Contract Level Risk Adjustment Data Validation Audits with an accompanying Technical Appendix. True and correct copy of these documents are attached hereto as Exhibits 1-3.[1] The proposed revisions to the regulations can also be found at https://s3.amazonaws.com/public-inspection.federalregister.gov/2018-23599.pdf and, starting November 1, 2018, at https://federalregister.gov/d/2018-23599. The Executive Summary and Technical Paper are at https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-Risk-Adjustment-Data-Validation-Program/Resources.html.

In the proposed revisions to the regulations, CMS proposes to extrapolate the results of its RADV audits to the contract-level without applying an FFS Adjuster. The reasons for not applying an adjuster are explained in its comments to the proposed regulatory revisions as well in the Executive Summary and its Technical Appendix. *See, e.g.*, Executive Summary (Exhibit 2) at pages 5-6. HHS' conclusions and findings relating to the FFS Adjuster issue are relevant to the "actuarial equivalence" and other issues currently before this Court due to the United States' pending Motion for Partial Summary Judgment (Dkt. Nos. 234-1, 250, and 272) and Motion Pursuant to Rule 16(b)(3) for Issuance of Scheduling Order (Dkt. Nos. 261-1, 274, and 278), which were heard on September 17, 2018. If the Court so orders, the United States welcomes the opportunity to file a short supplement brief explaining the importance of HHS'

[1] The proposed regulations address a number of unrelated issues and the entire document is 362 pages. Exhibit 1 consists of pages relating to the proposed revisions to the RADV audit regulations and the FFS Adjuster issue (pages 1, 9, 10, 13, 195-211, 236, 279-285, and 316-317). Exhibit 2 consists of the FFS Adjuster Executive Summary. Exhibit 3 consists of the FFS Adjuster Technical Appendix.

conclusions and findings to the Court's resolution of these motions.

Dated: October 29, 2018

Respectfully submitted,

JOSEPH H. HUNT
Assistant Attorney General, Civil Division
NICOLA T. HANNA
United States Attorney
DAVID M. HARRIS
DAVID K. BARRETT
ABRAHAM C. MELTZER
Assistant United States Attorneys

MICHAEL D. GRANSTON
DANIEL R. ANDERSON
EDWARD C. CROOKE
CAROL L. WALLACK
JUSTIN DRAYCOTT
PAUL G. FREEBORNE
JESSICA KRIEG
ZOILA E. HINSON
AMY L. LIKOFF
ADAM R. TAROSKY
Civil Division, Department of Justice

JAMES P. KENNEDY, JR.
United States Attorney
KATHLEEN ANN LYNCH
Assistant United States Attorney

/S/ John E. Lee

JOHN E. LEE, Assistant U.S. Attorney
Attorneys for the United States of America

Billing Code 4120-01-P

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Centers for Medicare & Medicaid Services

42 CFR Parts 422, 423, 438, and 498

[CMS-4185-P]

RIN 0938-AT59

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

AGENCY: Centers for Medicare & Medicaid Services (CMS), HHS.

ACTION: Proposed rule.

SUMMARY: This proposed rule would revise the Medicare Advantage (MA) program (Part C) regulations and Prescription Drug Benefit program (Part D) regulations to implement certain provisions of the Bipartisan Budget Act of 2018; improve quality and accessibility; clarify certain program integrity policies; reduce burden on providers, MA plans, and Part D sponsors through providing additional policy clarification; and implement other technical changes regarding quality improvement. This proposed rule would also revise the appeals and grievances requirements for Medicaid managed care and MA special needs plans for dually eligible individuals to implement certain provisions of the Bipartisan Budget Act of 2018.

DATES: To be assured consideration, comments must be received at one of the addresses provided below, no later than 5 p.m. on [insert date 60 days from date of publication in the Federal Register].

4. Preclusion List Requirements for Prescribers in Part D and Individuals and Entities in MA, Cost Plans, and PACE (§§ 422.222 and 423.120(c)(6))

In the April 2018 final rule, CMS removed several requirements pertaining to MA and Part D provider and prescriber enrollment that were to become effective on January 1, 2019. We stated in that final rule our belief that the best means of reducing the burden of the MA and Part D provider and prescriber enrollment requirements without compromising our payment safeguard objectives would be to focus on providers and prescribers that pose an elevated risk to Medicare beneficiaries and the Trust Funds. That is, rather than require the enrollment of MA providers and Part D prescribers regardless of the level of risk they might pose, we would prevent payment for MA items or services and Part D drugs that are, as applicable, furnished or prescribed by demonstrably problematic providers and prescribers. Therefore, we established in the April 2018 final rule a policy under which: (1) such problematic parties would be placed on a "preclusion list"; and (2) payment for MA services and items and Part D drugs furnished or prescribed by these individuals and entities would be rejected or denied, as applicable. The MA and Part D enrollment requirements, in short, were replaced with the payment-oriented approach of the preclusion list.

This proposed rule would make several revisions and additions to the preclusion list provisions we finalized in the April 2018 final rule. We believe these changes would help clarify for stakeholders CMS' expectations with respect to the preclusion list.

5. Medicare Advantage Risk Adjustment Data Validation (RADV) Provisions (§§ 422.300, 422.310(e), and 422.311(a))

The Medicare Advantage Risk Adjustment Data Validation (RADV) program was implemented as the primary corrective action to reduce the Part C improper payment rate in

compliance with the Improper Payments Information Act (IPIA) of 2002, as amended by the Improper Payments Elimination and Recovery Act (IPERA) of 2010 and updated by the Improper Payments Elimination and Recovery Improvement Act (IPERIA) of 2012. In this proposed rule, we would, based on longstanding case law and best practices from HHS and other federal agencies, establish that extrapolation may be utilized as a valid part of audit authority in Part C, as it has been historically a normal part of auditing practice throughout the Medicare program.

Accordingly, we are proposing the following:

• To establish that CMS would use extrapolation in RADV contract-level audits and that the extrapolation authority would apply to the payment year 2011 contract-level audits and all subsequent audits.

• Not to apply a fee-for-service (FFS) Adjuster to audit findings.

Provision	Description	Impact
MA Risk Adjustment Data Validation Provisions (§§ 422.300, 422.310(e), and 422.311(a))	We are proposing to establish that CMS would use extrapolation in RADV contract-level audits and that the extrapolation authority would apply to the payment year 2011 contract-level audits and all subsequent audits, and not to apply a FFS Adjuster to audit findings.	The estimated savings is $1 billion in 2020 to the Trust Fund, due to collections from industry of money improperly paid. In subsequent years the provision would save the Trust Fund at least $381 million each year. The savings result from the Trust Fund not making improper payments. Extrapolating audit findings does not increase the cost burden on the plan. The cost to the plan of complying with a RADV audit is neither the subject of nor affected by this provision. This provision addresses recovering extrapolated or non-extrapolated audit findings. While extrapolation does increase the level of the audit recovery, because returning improper payments is not a cost, the decision to extrapolate does not impact the cost to the plan.

2. Medicare Advantage Risk Adjustment Data Validation Provisions (§§ 422.300, 422.310(e), and 422.311(a))

a. Background

Subpart G of the MA regulations at part 422 describes how payment is made to MA organizations. These payment principles are based on sections 1853, 1854, and 1858 of the Act. Subpart G also sets forth the requirements for making payments to MA organizations offering local and regional MA plans, including calculation of MA capitation rates.

Section 1853(a)(3) of the Act requires that we risk adjust our payments to MA organizations. Risk adjustment strengthens the Medicare program by ensuring that accurate payments are made to MA organizations based on the health status plus demographic characteristics of their enrolled beneficiaries and ensures that MA organizations are paid appropriately for their plan enrollees (that is, less for healthier enrollees expected to incur lower health care costs and more for less healthy enrollees expected to incur higher health care costs). Accurate payments to MA organizations also help ensure that providers are paid appropriately for the services they provide to MA beneficiaries. In general, the current risk adjustment methodology relies on enrollee diagnoses and encounters, as specified by the International Classification of Disease, currently the Tenth Revision Clinical Modification guidelines (ICD-10-CM), to prospectively adjust capitation payments for a given enrollee based on the health status of the enrollee. Diagnosis codes determine the risk scores, which in turn determine the risk-adjusted payments. As a result, MA organizations and providers must focus attention on complete, truthful, and accurate diagnosis reporting according to the official ICD-10-CM coding guidelines.

As the ICD-10-CM guidelines emphasize, "accurate coding cannot be achieved" without "consistent, complete documentation in the medical record." Diagnoses submitted for payment by MA organizations must be supported by medical record documentation. This requirement has been in place since the beginning of the MA program. It has been explained in every edition of the Medicare Managed Care Manual, with which MA organizations agree to comply as a condition of their participation. (See the 2013 Medicare Managed Care Manual, § 40; 2004 Medicare Managed Care Manual, § 111.1, Ex. 30 & § 111.4; 2001 Medicare Managed Care Manual, § 110.4.) It has also been emphasized in numerous trainings provided to MA organizations and their subcontractors.

The diagnosis data submitted by MA organizations must conform to all relevant national standards. (See 42 CFR 422.310(d)(1).) As discussed earlier, the Clinical Modification of the International Classification of Disease, published by the federal government, is the chief national standard for diagnosis coding. It is the coding system on which MA risk adjustment is run. Medical record documentation is a core principle of the ICD-10-CM diagnosis coding system and was equally central to the Ninth Revision (ICD-9-CM), which preceded it. A federal court of appeals has recognized the requirement of medical record documentation for diagnosis codes submitted for payment by MA organizations. *United States ex rel. Swoben v. United Health Ins. Co.*, 848 F.3d 1161, 1168, 1176 (9th Cir. 2016). When MA organizations certify that their diagnosis codes are "accurate" and "truthful" to the "best knowledge, information, and belief" of the certifying individual, the existence of adequate medical record documentation is one important standard by which accuracy and truthfulness are measured (42 CFR 422.504(l)(1)). As we have previously explained, our "risk adjustment methodology provides that a specific amount be paid if an enrollee has a particular condition" (75 FR 19745). The medical record

documentation requirement is "designed to ensure that the enrollee in fact has th[e] condition" for which an MA organization is requesting payment under the risk adjustment model (75 FR 19745).

The current risk adjustment model employed in adjusting MA plan payments is known as the CMS Hierarchical Condition Category (CMS-HCC) model. It functions by categorizing ICD-10-CM codes into disease groups called Hierarchical Condition Categories, or HCCs. Each HCC includes diagnosis codes that are related clinically and have similar cost implications. The CMS-HCC model is recalibrated approximately every 2 years to reflect newer treatment and coding patterns in Medicare FFS. This recalibration is made through the annual advance notice of methodological changes authorized by 42 U.S.C. 1395w-23(b)(2). Since 2007, when a demographic data-only payment method was completely phased-out for MA plans, 100 percent of payment has been risk-adjusted. The statute continues to provide us the authority to add to, modify, or substitute for risk adjustment factors if the changes will improve the determination of actuarial equivalence.

b. Risk Adjustment Data Validation Initiatives

MA enrollee HCCs are assigned based on data submitted to us by MA organizations via the Risk Adjustment Payment System (RAPS) and Encounter Data System (EDS). The HCCs contribute to an enrollee's risk score, which is used to adjust a base payment rate. Essentially, the higher the risk score for an enrollee, the higher the expected health care cost for the enrollee. The HCC data that MA organizations submit to CMS via the RAPS and EDS systems is self-reported by the MA organization and does not go through a validation review before being incorporated into a given beneficiary's risk-profile. Since there is an incentive for MA organizations to potentially over-report diagnoses so that they can increase their payment, the

Department audits plan-submitted diagnosis data a few years later to ensure they are supported by medical record documentation.

Verifiable medical record documentation is key to accurate payment and successful data validation. We annually select MA organizations for risk adjustment data validation (RADV) audits.[23] RADV audits are intended to confirm the presence of risk adjustment conditions (that is, diagnoses that map to HCCs) as reported by MA organizations for their enrollees and confirmed via medical record documentation. RADV audits occur after the final risk adjustment data submission deadline for the MA contract year. The audits validate the HCC data submitted by MA organizations by reviewing hospital inpatient, hospital outpatient, and physician/practitioner provider medical records. The focus of this medical record review activity is on diagnoses related to the enrollee's HCC profile. Risk adjustment discrepancies are identified when the enrollee's HCCs used for payment (based upon MA organization-submitted data) differ from the HCCs assigned based on the medical record, pursuant to the RADV audit process. Risk adjustment discrepancies can be aggregated to determine an overall level of payment error. In turn, payment error for a sample of contract enrollees can be extrapolated to calculate a contract-level payment error estimate. Although we have the authority to extrapolate from a statistically valid sample to calculate a contract-level audit recovery, we have not yet done so.

From 1999 until 2003, our payment validation activity for the MA program had both an educational and audit focus and was intended to improve the accuracy of the risk adjustment data that was being submitted to CMS for payment. Payment adjustments were limited to enrollee-level adjustments for those enrollees sampled in the payment validation audit. At the time, only

[23] Any changes to the CMS-HCC payment model are published in the annual payment notice.

10 percent of the MA payment amount was risk adjusted. As a result, payment recovery amounts for the small number of plans audited was very small. Since payment year 2004 was the first year for which MA payments were based on the current HCC risk adjustment model, we considered payment years 2004 through 2006 as pilot years for the purpose of RADV and no payment recovery activity occurred.

Payment recovery resumed for payment year 2007, when we audited 37 MA contracts and recouped $13.7 million. Payment adjustments were again limited to enrollee-level adjustments for those enrollees sampled in the payment validation audit. (Although we suggested that we would make contract-level payment adjustments for the payment year 2007 audits, we did not ultimately do so.) In the course of that audit process, as in previous years, we reviewed medical record documentation provided by each audited MA organization to substantiate conditions reported by the organization for beneficiaries in each audit sample. After CMS' findings were reported to each MA organization, any organization that disagreed with CMS' determinations could challenge them through a three-stage administrative process established by regulation in 2010. (See 42 CFR 422.311). This dispute and appeals process is currently ongoing.

No payment validation audits were conducted for payment years 2008, 2009, or 2010. In those years, we were considering the development of a methodology for calculating payment adjustments based on statistical RADV MA contract-level payment error audit findings. The development of contract-level RADV audits would enable us to make contract-level payment adjustments rather than simply adjusting payments for specific enrollees from an audit sample, as we had done previously.

On December 20, 2010, we proposed a methodology on the CMS website for selecting a statistically-valid sample of enrollees from each audited MA contract and extrapolating from the results of that sample audit to calculate a contract-level payment adjustment. We invited public comment on this proposed methodology, and received more than 500 comments, which we carefully reviewed. On February 24, 2012, we published what we described as the final methodology for RADV contract-level payment error calculation.[24] That methodology described sampling techniques and the statistical calculation to be used to extrapolate from the sample selected. In brief, up to 201 enrollees from each audited MA contract would be selected according to certain criteria, including their continuous enrollment in the contract for the entire data collection year and January of the payment year; their lack of end-stage renal disease (ESRD) status and hospice status for that entire period; their enrollment in Medicare Part B coverage for the entire data collection year; and their submission of at least one diagnosis during the data collection year leading to at least one CMS-HCC assignment in the payment year. The RADV-eligible enrollees would be ranked by risk score and then divided into three equal strata. An equal number of enrollees would then be randomly selected from each stratum (67 enrollees per stratum in the case of an audit of 201 enrollees). After medical records were reviewed, payment errors would be calculated for each selected enrollee based on the number of months the person was enrolled in the selected MA contract (and was not in ESRD or hospice status) during the payment year. A payment error rate for each stratum would be calculated, and then an overall payment error rate for the audited contract, computed at a ninety-nine percent confidence interval. We stated that this methodology would be applied to the next round of RADV audits,

[24] 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.

which would be conducted on payment year 2011. Audits for payment years 2011, 2012, and 2013 have been conducted according to this methodology, at a total cost of approximately $150 million to the agency, but have not yet been finalized. These audits are in addition to RADV and related MA audits conducted by the Office of Inspector General, which are conducted pursuant to OIG's independent authorities at sections 2(1) and 4(a)(1) of the Inspector General Act.

We also stated in 2012 that, after using this methodology to calculate a preliminary payment recovery amount, we would apply a FFS Adjuster as an offset before finalizing the audit recovery. The FFS Adjuster was intended to account for any effect of erroneous diagnosis codes in the data from Medicare Parts A and B (often referred to as "Fee-For-Service" Medicare) that are used to calibrate the MA risk adjustment model. We stated that the FFS Adjuster would calculate a permissible level of payment error (for example, a percentage of the total payments made on an MA contract in a given year) and limit RADV audit recovery to payment errors above that level. The FFS Adjuster was never intended to set a permissible rate for the submission of erroneous diagnosis codes. We stated that the FFS Adjuster would be calculated based on a RADV-like review of records submitted to support the Medicare Part A and B diagnosis codes. That review is now complete, and will be discussed later.

c. Discussion of Proposals

(1) Extrapolation

The Secretary intends to recover overpayments based on extrapolated audit findings through the use of statistically valid random sampling techniques. Although we described our February 2012 publication as the final methodology to be used to calculate contract-level RADV audit recoveries for payment year 2011, it has never been implemented. As we stated earlier, audits for payment years 2011, 2012, and 2013 have been conducted according to this

methodology, but contract-level recoveries have not yet been sought. We are now providing additional notice and again welcoming public input on the agency's methodology for calculating a contract-level payment error in RADV audits, including the sample sizes used in these contract-level audits. CMS is not required to set forth the methodology for calculating an extrapolated payment error through regulatory provisions (it does not do so in Parts A and B, where Medicare Administrative Contractors (MACs) may use any statistically valid sampling and extrapolation methodology they determine to be appropriate), however, in the interest of transparency, we are updating stakeholders on our plans to use various sampling and extrapolation methodologies in RADV audits, as CMS deems appropriate.[25] All audits will be based on statistically valid sampling and extrapolation methodologies.

In addition to the contract-level methodology described earlier, we have identified other potential methodologies for sampling and extrapolation, which would calculate improper payments made on the audited MA contract for a particular sub-cohort or sub-cohorts in a given payment year, and the agency may also use such a methodology to calculate improper payments made to the audited MA contract. For example, a sub-cohort could be the enrollees for whom a particular HCC or one of a related set of HCCs (such as the three diabetes HCCs) was reported. After choosing an MA contract and a sub-cohort or sub-cohorts to audit, we would select a statistically significant sample of enrollees for the sub-cohort or sub-cohorts. After reviewing the medical records of those enrollees, we would use statistical extrapolation to calculate and recoup the improper payments made to the audited MA contract for covering enrollees for the sub-cohort or sub-cohorts in that payment year. We would use the same statistical calculation

[25] The Office of the Inspector General, which is required by law to conduct audits and follow generally accepted government auditing standards, does not seek comment on its methodology for risk adjustment audit work that may lead to overpayment recoveries from MA organizations.

for this sub-cohort-level extrapolation as we do for the contract-level extrapolation (although we welcome comment as to whether to stratify the sample population for the sub-cohort audits, as we currently anticipate doing for the contract-level audits).

We believe that, because any sub-cohort is necessarily a subset of the enrollees covered through a particular MA contract, we could often use a much smaller sample size to calculate a statistically significant extrapolated recovery for a sub-cohort than would be required to calculate a contract-level recovery (up to 201 enrollees, according to our anticipated contract-level methodology). This smaller sample size would allow us to spread our audit resources across a wider range of MA contracts, while still generating statistically significant recoveries. This sub-cohort-based audit methodology would allow us to focus on cohorts of enrollees that appear to raise programmatic concerns.

We invite comment on both the contract-level audit methodology published in February 2012, and our proposal for an extrapolated audit methodology based on sub-cohorts of enrollees. We also seek comment on whether there are particular situations in which one methodology may be preferable to the other, and whether the agency should revise the contract-level audits that have been conducted but not finalized for payment years 2011, 2012, and 2013. Neither proposed methodology is meant to displace our longstanding authority to audit the medical records of particular enrollees who we believe may be associated with improper payments or to use any statistically valid audit methodology. [26]

[26] We may begin to conduct RADV audits for payment years 2014 and 2015 before this proposal is finalized, pursuant to our longstanding authority to review the medical records of any MA enrollee and recoup any improper payments identified. Although we would design these audits so that that the individuals selected would form a statistically significant sample that would support an extrapolated recovery, we would not seek to recover on an extrapolated basis until the rule is final. At the very least, these audits would support enrollee-level recoveries.

If we finalize one or more sampling and extrapolation methodologies through this rulemaking, we would make any future changes to that methodology (or those methodologies) through the Health Plan Management System.

We are also considering whether to explicitly expand the MA organizations' RADV appeal rights, particularly in light of the upcoming auditing and recoveries in the MA program. One option would be to permit appeal of the RADV payment error calculation methodology used in a RADV audit similar to practices in the Part A and Part B space of Medicare FFS. We invite comments on this matter.

(2) Application to Payment Year 2011 and Subsequent Years

We intend to apply the finalized RADV payment error methodology or methodologies to payment year 2011, and all subsequent years. (However, we do not expect to use a sub-cohort-based methodology, if finalized, for any payment year before 2014.) Section 1871(e)(1)(A) of the Act authorizes retroactive application of rules where "(i) such retroactive application is necessary to comply with statutory requirements; or (ii) failure to apply the change would be contrary to the public interest." We are considering whether application of the finalized methodology or methodologies to payment year 2011, and all subsequent years, would require the exercise of this statutory authority to engage in retroactive rulemaking. We invite comment on the subject.

In any case, we believe that failure to apply the finalized RADV payment error methodology or methodologies to those payment years would be contrary to the public interest. The public has a substantial interest in the recoupment of millions of dollars of public money improperly paid to private insurers. The public also has a significant interest in providing incentives for those insurers to claim only proper payments in the future, which would be

promoted by the recoupment of funds improperly paid in the past. Given the amount of improper payments identified under the MA program (estimated to be $14.35 billion in FY 2017,[27] the $650 million in recovered improper payments represents, if this policy was finalized, 3 years improper payment for 30 plans), the interest in determining an accurate recovery amount for each audited MA plan, and the importance of protecting the overall integrity of the program, we believe that it is in the public interest for CMS to apply the RADV payment error methodology or methodologies adopted through this rulemaking to payment year 2011 and all subsequent years. In applying this methodology (or these methodologies) to those payment years, CMS would be acting in compliance with the IPERIA statute[28] as well as its own fiduciary responsibility to recover funds due and owing to the Medicare Trust Funds. We note also that our February 2012 publication put MA organizations on notice that CMS expected to calculate a contract-level payment error for payment year 2011 and beyond by extrapolating from its review of a statistically valid sample of enrollees, and that (as explained earlier) MA organizations have never been entitled to receive or retain payments associated with HCCs that cannot be validated

[27] CMS has historically reported high levels of payment error in the Part C program. The Part C error rate has ranged between 11 percent and 9 percent between fiscal years (FY) 2011 and 2014, respectively. In FY 2017, the reported Part C error rate was 8.31 percent or $14.35 billion.

[28] Improper Payments Elimination and Recovery Improvement Act of 2012 (IPERIA, Pub. L. 112-248). The RADV program is a corrective audit activity developed by CMS to address provisions included in the IPIA of 2002, as amended by the IPERA of 2010, and further amended by IPERIA. These statutes require that government agencies annually estimate and report improper payments. RADV audits were initiated because Part C payment error was out of compliance with IPIA. The IPERIA requires the Office of Management and Budget (OMB) to annually identify agencies for greater levels of oversight and review, and with that agency "establish annual targets and semi-annual or quarterly actions for reducing improper payments associated with each high-priority program." In November 2009, Executive Order (E.O.) 13520 was signed in an effort to reduce improper payments by increasing transparency in government and holding agencies accountable for reducing improper payments. In March 2010, OMB issued guidance for agencies regarding the implementation of E.O. 13520 entitled Part III to OMB Circular A-123, Appendix C (Appendix C). Appendix C outlines the responsibilities of agencies, determines the programs subject to E.O. 13520, defines supplemental measures and targets for high priority programs, and establishes reporting requirements under E.O. 13520 and procedures to identify entities with outstanding payments. One of those remedies is payment recapture audits, a requirement that any program that expends at least $1million must implement payment recapture audits. A recovery audit, or payment recapture audit, is a review process designed to identify erroneous payments. Additionally, it is a corrective control activity designed to identify and recapture erroneous payments, and, as such, is a management function and responsibility.

by medical records. Application of the finalized RADV payment error methodology or methodologies to payment year 2011 and all subsequent years therefore would not upset any settled interest.

If the finalized contract-level audit methodology differs from the one we published in February 2012, we will also consider whether to apply the new contract-level payment error methodology to payment years 2011, 2012, and 2013, or to only apply it to payment year 2014 and subsequent years, and to finalize the audits for those earlier payment years according to the methodology published in February 2012. We invite comment on this subject, as well. In any event, and however audits for prior years are ultimately handled, we believe that it is vitally important for the health of the MA program to have extrapolated recoveries available for future audit years.

(3) Implementation

This proposal would announce CMS' intention to recover improper payments based on extrapolation of payment error from RADV audit samples to MA organization specified populations. CMS would calculate and recover improper payments based on extrapolation methodologies. MA organizations would be required to remit extrapolated recovery amounts from audit findings as calculated by CMS through its payment system, Medicare Advantage and Prescription Drug system (MARx). MARx is the CMS system that makes monthly payments and payment adjustments to the MA organizations and Part D sponsors. Overpayment recoveries of all types are considered payment adjustments which are done as offsets to the plans' monthly payments. RADV recovery amounts are included in this category. In the month the plan has been notified that the recovery amount will be offset, the MARx system makes an offset to the plans monthly payment equal to the amount of the recovery amount. In the event the recovery amount

exceeds the payment in 1 month, the recovery will be spread across adjustments for multiple months until the full amount is recovered. CMS may likewise require MA organizations to remit such recovery amounts based upon audit findings by OIG.

(4) Recoupment of Improper Payments in Part C

Improper payments identified by CMS outside of the RADV audit process or self-identified by the MA organization that are not returned in accordance with §§ 422.330, and are identified and/or estimated through extrapolation or other estimation methodologies as a result of CMS audits will be recovered following CMS audit processes including payment offset. We propose that MA organizations be required to remit funds that CMS calculates as improper payments through the extrapolated RADV audit findings in accordance with §§ 422.310(e). RADV audit results can be appealed by MA organizations using the regulatory administrative appeals process outlined in § 422.311.

(5) FFS Adjuster

After our 2012 RADV publication, we conducted an extensive study regarding the presence and impact of diagnosis error in FFS claims data. Our study suggests 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.[29]

The study began by auditing 8,630 outpatient claims paid through Medicare Part B in a given year. We reviewed the medical records associated with each claim (a small subset of the medical records associated with each beneficiary) to determine whether the diagnosis associated

[29] 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.

with the claim was supported by medical record documentation. A discrepancy rate for each CMS-HCC was then calculated. For example, the data set contained 484 claims submitted with a diagnosis of chronic obstructive pulmonary disease, which is CMS-HCC 108. Of those diagnoses, 388 were supported by medical record documentation, and 96 were not, for a discrepancy rate of 19.8 percent. To account for the fact that the data set contained extremely small samples of many CMS-HCCs—for example, one diagnosis of extensive third degree burns and two diagnoses of severe head injury—we calculated a high, low, and baseline discrepancy rate. Each CMS-HCC was assigned one of these three mean discrepancy rates depending on its relationship to the bassline discrepancy rate: CMS-HCCs with a discrepancy rate significantly higher than the baseline were assigned to the high category, and those with a discrepancy rate significantly lower than the baseline were assigned to the low category. All other CMS-HCCs were assigned the baseline discrepancy rate. These rates were 46.2 percent, 33.8 percent, and 20.9 percent.

In a given year, multiple claims are submitted for Medicare Part B services received by a given beneficiary and associated with a given diagnosis. For example, an average beneficiary with metastatic cancer or acute leukemia, which is CMS-HCC 7, has seven claims associated with that diagnosis. Because we were interested in determining whether a given beneficiary had a documented diagnosis in a given year, and not whether any particular claim was associated with medical record documentation, we used the claim-level discrepancy rates described above to calculate beneficiary-level discrepancy rates.[30]

[30] For example, metastatic cancer or acute leukemia was assigned the baseline discrepancy rate of 33.8%. We therefore reasoned that each of the seven claims associated with the average beneficiary for whom such a diagnosis was reported had a 66.2% chance of being supported by medical record documentation, and only one instance of medical record support was necessary to make the diagnosis valid for that year. If each beneficiary with such a reported diagnosis has 7 claims associated with that diagnosis, and each claim has a 66.2% chance of being

After calculating this beneficiary-level discrepancy rate for each HCC, we ran fifty simulations in which we removed diagnoses from a data set of more than 1.4 million Medicare Part A and B beneficiaries at the beneficiary-level discrepancy rate.[31] After removing diagnoses at the indicated rates, we used each simulated "corrected" data set to recalibrate the CMS-HCC risk adjustment model, applied the recalibrated risk coefficients to a data set of MA beneficiaries, and compared their original risk scores to the risk scores calculated with the recalibrated model. We found that the difference between the risk scores was very small, and that the recalibrated risk scores tended to be slightly lower than the original risk scores. Therefore, we concluded that diagnosis error in FFS claims data does not lead to systematic payment error in the MA program.

An executive summary of the findings and a technical appendix describing the data and methodology can be found at https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-Risk-Adjustment-Data-Validation-Program/Resources.html. Because it appears that diagnosis error in FFS claims data does not lead to systematic payment error in the MA program, we propose not to include an FFS Adjuster in any final RADV payment error methodology.

Moreover, even if we 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. RADV audits are used to recover payments based on diagnoses that are not supported by medical record documentation, which thus should not have been reported to CMS. If a payment has been made to an MA organization based on a diagnosis

---

supported by medical record documentation, then 99.95% of all beneficiaries will have at least one instance of medical record support, and only 0.05% of beneficiaries will lack any medical record documentation of their reported diagnosis.

[31] For metastatic cancer and acute leukemia, 1 in 2,000 diagnoses was removed (corresponding to an error rate of 0.05%).

code that is not supported by medical record documentation, that entire payment is in error and should be recovered in full, because the payment standard has not been met, and the MA organization is not entitled to any payment for that diagnosis. RADV audits do not address issues with the accuracy of payments based on diagnosis codes that are supported by medical record documentation. Consequently, an adjustment to RADV recoveries to remedy payment accuracy concerns is inappropriate. For this reason, we believe that it would not be appropriate to correct any systematic payment error in the MA program through a payment adjustment that was only applied to audited contracts. Doing so would introduce inequities between audited and unaudited plans, by only correcting the payments made to audited plans.

Because our study suggests that diagnosis error in FFS claims data does not lead to systematic payment error in the MA program and because we believe it would be inequitable to correct any systematic errors in the payments made to audited plans only, we would not include an FFS Adjuster in any RADV extrapolated audit methodology. We welcome public comments on this study.

d. Proposed Changes

In this section, we discuss the proposed changes to the regulation in Parts 422 and 423 governing the MA Program. We are proposing to apply extrapolation to plan year audits for payment year 2011 forward.

The following is a summary of the proposed changes included in this proposed revision:

We propose to revise § 422.300 to include "collection of improper payments."

We propose to amend § 422.310(e) Validation of risk adjustment data, to apply extrapolation to plan year audits for payment year 2011 forward.

We propose to amend § 422.310(e) Validation of risk adjustment data, by adding a requirement to set forth the provision for MA organizations to remit improper payments based on RADV audits and established in accordance with stated methodology, in a manner specified by CMS.

We propose to amend § 422.311, the RADV audit dispute and appeal process section, by adding language to clarify that recovery of improper payments from MA organizations will be conducted according to the Secretary's payment error extrapolation and recovery methodologies and that CMS will apply extrapolation to plan year audits for payment year 2011 forward.

policy and would not impose any new or revised information collection requirements (that is, reporting, recordkeeping, or third-party disclosure requirements) or burden. Consequently, the provisions are not subject to the PRA.

8. ICRs Regarding Medicare Advantage Risk Adjustment Data Validation Provisions (§§ 422.300, 422.310(e), and 422.311(a))

As described in section III.C.2. of this proposed rule, we are proposing that extrapolation may be utilized as a valid part of audit authority in Part C, as it has been historically a normal part of auditing practice throughout the Medicare program. We are also proposing that this extrapolation authority be applied to the payment year 2011 RADV contract-level audits and all subsequent audits to reduce the Part C improper payment rate. Additionally, we are proposing not to apply a FFS Adjuster to audit findings.

The proposed provisions would not impose any new or revised information collection requirements (that is, reporting, recordkeeping, or third-party disclosure requirements) or burden since the utilization of extrapolation will not affect the existing process for MA organizations submitting medical record documentation pursuant to RADV audits. Consequently, the provisions are not subject to the PRA.

described in the April 2018 final rule. Our proposed provisions are, generally speaking, clarifications of our intended policy and do not constitute new requirements. Hence, the expected impact is negligible.

8. Medicare Advantage Risk Adjustment Data Validation Provisions (§§ 422.300, 422.310(e), and 422.311(a))

a. Proposals

This proposed rule would create regulations to govern the collection of *extrapolated* audit findings. As CMS develops its approach to statistical sampling and extrapolation, it is taking account of the recommendations of the 2016 General Accounting Office (GAO) report on CMS audit practices.[52] For example, CMS has been randomly selecting 30 plans for audit based on factors unrelated to payment error. In recent years, only half of those audited plans have had findings; the other half have had no net findings of improper payments. The GAO has recommended that CMS select plans that historically have high error rates either from the National audits as published in the Report of the Chief Financial Officer or from prior CMS audits. This recommendation would probably increase the number of findings, and hence the amount collected through the audits. CMS has accepted all GAO findings and intends to develop its sampling and extrapolation methodology consistent with them.

To clarify in more detail how the proposed rules would impact the recovery audit process we note the following facts:

- RADV recovery for payment years 2011, 2012, and 2013 included 30 MA contracts per payment year. For each contract, 200 enrollees have been selected. The aggregate cost to the government for each audit is $54 million.

[52] https://www.gao.gov/products/GAO-16-76

- National audits are for the purpose of payment error measurement in the Part C program. A nationally representative sample of 600 enrollees are selected from approximately 200 plans. Each plan contributes between 1 to 15 enrollees with many plans contributing under 10 enrollees. The annual cost to the government of a national audit is between $6 to 10 million. No recovery is made through the national audits.

- Findings from the national and contract-level audits will be used to predict beneficiaries at most risk for improper payment. CMS will use these estimates to target plans at most risk for improper payment for RADV audit.

- By better targeting audits to improper payment, CMS expects any sentinel effect of RADV to continue to reduce the historical Part C error rate.

b. Expected Impact of These Provisions

While we cannot fully estimate the quantitative impact of this provision, we can clearly identify certain components of impact. We start with some basic facts mentioned in the preceding narrative.

- With extrapolated audit findings, we would realize a positive ROI. The cost per year for a RADV audit is $54 million. Non-extrapolated recoveries would result in a $10 to 15 million collection per audit.

- Extrapolating audit findings does not increase the cost burden on the plan. The cost to the plan of complying with a RADV audit is neither the subject of nor affected by this provision. This provision addresses recovering extrapolated or non-extrapolated audit findings. While extrapolation does increase the level of the audit recovery, because returning improper payments is not a cost, the decision to extrapolate does not impact the cost to the plan.

- The audits for payment years 2011, 2012, and 2013 suggest that audited MA contracts received $650 million in of improper payments in those 3 years.
- This $650 million would be a transfer from the government to insurers since money paid for human coding error which CMS paid the contracts to pay their providers is no longer being done, meaning that the contracts must take responsibility for the improper provider payments.
- These audits cover 3 years, with 30 contracts audited each year.
- Roughly half the contracts each year had no net findings of improper payments.

Using these data we can conclude as follows:

- The audits for payment years 2011, 2012, and 2013 suggest that audited MA contracts were responsible for $650 million of improper payments in those 3 years.
- $650 million divided by 3 audit years is $217 million per audit year.
- $217 million per audit year divided by 15 contracts with audit findings per year is $14.5 million per contract with audit findings per year.
- If GAO recommendations are adopted which would facilitate focusing on contracts with expected findings, and the level of audit findings holds constant, then $14.5 million per contract with audit findings per year times 30 contract with audit findings per year would produce $435 million in audit collections per year.
- This level of recovery would produce $381 million in aggregate savings per year (that is, $435 million - $54 million, since the cost of audits would remain at $54 million).

This numerical bulleted argument is summarized in Table 15.

It might seem natural to trend the $381 million based on non-inflation factors. The following considerations argue against trending. Therefore, we are leaving the estimate of dollar savings to

the Medicare Trust Fund at $381 million per year at each year for the next 10 years with an additional $650 million the first year. A 10-year table is presented in Table 16. The arguments against trending are the following:

- The error rate of improper payments per year, as indicated in the reports of the Chief Financial Officer have been declining and are likely to continue to decline. Importantly, although we have about 10 years of data we have insufficient data to extrapolate since performance error is rarely linear. Thus trending would involve non-linear functions and would require more data.

- The aggregate amount paid to contracts is increasing due to enrollment growth. The Office of the Actuary at CMS annually publishes a Trustee Report which contains projected enrollment.[53]

- The $381 million is based on current error rates and enrollment growth. But we have already indicated that 50 percent of contracts audited had no net audit findings. We have already indicated that acceptance of GAO recommendations would facilitate targeting contracts with higher rates and have therefore assumed there would be findings in all 30 contracts audited.

For these reasons, we are leaving the annual estimate as a dollar savings to the Medicare Trust Fund of $381 million for 2021 and future years, and a dollar savings of $1.03 billion to the Medicare Trust Fund in 2020 ($381 million savings per year plus an estimated $650 million in audit recoveries for payment years 2011 through 2013). All other things being equal, the increase in enrollment will cause the nominal dollars in error to increase. The historical decline in the error rate may or may not offset the increase due to increasing enrollment making a

[53] https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/ReportsTrustFunds/index.html

projection difficult. For this reason we hold the estimate of $381 million constant in the projection.

A table of collection for 10 years is summarized in Table 16.

TABLE 15: EXPECTED SAVINGS PER YEAR FROM RADV PROVISION

Label	Item	Amount ($ in millions)	Source or Calculation
(A)	Estimated Collection 2011-2013	$650	
(B)	Number of years, 2011-2013	3	
(C)	Estimated Collection per year, 2011-2013	$217	(C) = (A)/(B)
(D)	Number of contracts audited	30	
(E)	Percent of contracts with findings	50%	
(F)	Current Number of contracts with findings	15	(F) = (D)*(E)
(G)	Estimated Collection per year per contract	$14.5	(G)=(C)/(F)
(H)	Expected number of contracts with findings	30	If GAO report recommendations are adopted
(I)	Estimated collection per year	$435	(I)=(G)*(H)
(J)	Audit Cost per year	$54	Constant cost of auditing 200 beneficiaries per contract.
(K)	Estimated savings per year	$381	(K)= (I) - (J)

TABLE 16: IMPACT PER YEAR FROM RADV (IN MILLIONS)

	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029
Cost of Audit	(54)	(54)	(54)	(54)	(54)	(54)	(54)	(54)	(54)	(54)
Estimated Collection Prior Years	650	0	0	0	0	0	0	0	0	0
Estimated Collection This year	435	435	435	435	435	435	435	435	435	435
Estimated Total Savings	1031	381	381	381	381	381	381	381	381	381

The estimated 10-year dollar savings to the Medicare Trust Fund could be $4.5 billion ($381 million per year * 10 years + initial $650 million recovery).

The savings come from recovered inaccurate payments of $381 million a year by the Medicare Trust Fund to plans. This money is a reduction in spending of the Medicare Trust Fund (to the plans); there will be no money transferred to enrollees. We expect that ultimately this provision could incentivize plans to submit more accurate risk-adjustment data.

The intent of this rule is to continue the sentinel effect on the reduction of the Part C error rate. The decline in the Part C error rate has correlated with the announcement of the agencies intent to use extrapolated recoveries on payment years 2011 through 2013. We believe that forgoing the extrapolation on those audits would diminish the agency's credibility going forward and consequently reduce the sentinel effect. The dollar savings to the Medicare Trust Fund are presented in Table 16. In any case, RADV audits will still continue.

D. Alternatives Considered

1. Requirements for Medicare Advantage Plans Offering Additional Telehealth Benefits (§§ 422.100, 422.135, 422.252, 422.254, and 422.264)

Section 1852(m)(2)(A)(i) of the Act, as added by the Bipartisan Budget Act of 2018, defines additional telehealth benefits as services that are identified for the applicable year as clinically appropriate to furnish using electronic information and telecommunications technology when a physician (as defined in section 1861(r) of the Act) or practitioner (described in section 1842(b)(18)(C) of the Act) providing the service is not at the same location as the plan enrollee (which we refer to as "through electronic exchange"). We considered various alternative definitions of "clinically appropriate" but decided not to propose specific regulation text defining the term. We are proposing to implement the statutory requirement for additional telehealth

local MA plan, adjusted using the factors described in paragraph (c) of this section.

(3) The *risk adjusted MA region-specific non-drug monthly benchmark amount* is the unadjusted benchmark amount for coverage of basic benefits defined in § 422.100(c)(1) by a regional MA plan, adjusted using the factors described in paragraph (e) of this section.

* * * * *

17. Section 422.300 is revised to read as follows:

§ 422.300 Basis and scope.

This subpart is based on 42 U.S.C. 1106, 1128j(d), 1852, 1853, 1854, and 1858. It sets forth the rules for making payments to MA organizations offering local and regional MA policies, including calculation of MA capitation rates and benchmarks, conditions under which payment is based on plan bids, adjustments to capitation rates (including risk adjustment), collection of risk adjustment data, conditions for use and disclosure of risk adjustment data, collection of improper payments and other payment rules. See § 422.458 for rules on risk sharing payments to MA regional organizations.

18. Section 422.310 is amended by revising paragraph (e) to read as follows:

§ 422.310 Risk adjustment data.

* * * * *

(e) *Validation of risk adjustment data.* MA organizations and their providers and practitioners will be required to submit a sample of medical records for the validation of risk adjustment data, as required by CMS. There may be penalties for submission of false data. 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.

* * * * *

19. Section 422.311 is amended by revising paragraph (a) to read as follows:

§ 422.311 RADV audit dispute and appeal processes.

(a) *Risk adjustment data validation (RADV) audits.* In accordance with §§ 422.2 and 422.310(e), the Secretary annually conducts RADV audits to ensure risk adjusted payment integrity and accuracy. 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.

* * * * *

20. Section 422.504 is amended by adding paragraph (g)(1)(iv) to read as follows:

§ 422.504 Contract provisions.

* * * * *

(g) * * *

(1) * * *

(iv) The enrollee shall not have any financial liability for services or items furnished to the enrollee by an MA contracted individual or entity on the preclusion list, as defined in § 422.2 and as described in § 422.222.

* * * * *

21. Section 422.560 is amended by adding paragraphs (a)(4) and (b)(5) to read as follows:

§ 422.560 Basis and scope.

(a) * * *

October 26, 2018

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

I. Issue

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

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

II. Analytical Approach and Empirical Findings

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

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

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

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

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

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

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

---

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

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

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

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

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

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

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

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

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

5 A detailed explanation appears in technical appendix.

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

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

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

Table 1. Audit Miscalibration Estimate

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

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

6 A detailed explanation appears in technical appendix.

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

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

III. Conceptual Issues with the FFS Adjuster

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

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

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

IV. Conclusion

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

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

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

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

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

Development of the Fee for Service Adjuster (FFSA) was initiated as a response to the Medicare Advantage (MA) organizations' assertion that because the Centers for Medicare and Medicaid Services - Hierarchical Condition Category (CMS-HCC) risk adjustment model is calibrated on FFS diagnoses that are not validated by medical record documentation, there is an inaccuracy in payment that should be accounted for in Risk Adjustment Data Validation (RADV) audit recoveries. They have argued that the inaccuracy due to audit miscalibration introduces a systematic bias that causes underpayments to plans relative to FFS, which in turn is exacerbated by RADV payment error recoveries. [1]

In this section, we provide an expanded discussion on the theoretical, economic, and empirical grounds to evaluate the conception of the FFSA. The issues we focus on are as follows:

(1) Statistically, the foundations that would lead to such a payment bias are largely absent.
(2) Application of a FFSA to RADV recoveries is not an appropriate way to adjust for this bias without arbitrarily changing the payment methodology for plans deemed to owe recoveries vis-à-vis unaudited plans.
(3) Our empirical analysis does not provide evidence of a calibration error bias on the MA population.

I. Weak Statistical Foundations

The statistical foundation that would lead to a consistent bias is largely non-existent within the context of the risk adjustment modeling framework for MA. The discussion below examines how a bias can arise and why our circumstances are different.

Assume one wished to estimate an econometric model and apply it to the same population upon which the data was generated. In regression modeling, independent variables are assumed fixed and free of measurement error. If independent variables have measurement error, the affected regression estimates may be biased downward.

We first look at the econometric framework given the above assumptions. A simplified CMS-HCC model can be the following

$$Exp = \alpha D + \beta HCC^* + e$$

[1] In the context of this discussion, "underpayments" and "overpayments" are evaluated relative to the expected true FFS costs of a given model coefficient. They are not RADV overpayments and underpayments.

where Exp are FFS expenditures and HCC^* has the value of one if the sole HCC is present and zero if the HCC is not present and D represents the existence of a demographic characteristic.

We measure HCC^* with error so

$$HCC = HCC^* + w$$

Which in turn implies

$$HCC^* = HCC - w$$

Here, w is the measurement error of the HCC.

Thus, instead of estimating $Exp = \alpha D + \beta HCC^* + e$, we in fact estimate

$$Exp = \alpha D + \beta(HCC - w) + e$$

Rearranging terms, we can see this is

$$Exp = \alpha D + \beta(HCC) + (e - \beta w)$$

or

$$Exp = \alpha D + \beta(HCC) + z$$

Where the actual error term $z = (e - \beta w)$.

Under Ordinary Least Squares (OLS) assumptions, the error term is not correlated with the independent regressors. In this case, we see that z (the error term) contains the parameters of the HCC regressors. Thus, the error term is correlated with the regressors and therefore ostensibly violates the OLS assumptions. This may lead to bias and inconsistent estimates.

The central issue is the covariance between z and the HCC.

It can be shown that this covariance is

$$cov(z, HCC) = -\beta\sigma_w^2$$

Now we see that the extent of the problem depends on the variance of the measurement error from observation to observation of the FFS data. This is an empirical question. There are circumstances where there would be no significant bias.

In practice it may not be unreasonable to assume that the measurement error is fairly constant for the HCC across FFS observations. For example, the probability of error of the HCC may be ten percent. Under this assumption, the variance would be close to zero (the variance of a constant is zero) and by extension the covariance would also be close to zero. This in turn implies that the OLS assumptions essentially hold and the regression

procedure produces unbiased and consistent estimates. The variance may or may not be too small to have a substantial impact. The essential point is that, even if the problem conceptually exists, it may not have much effect in practice. The impact must be measurable, it cannot simply be assumed to exist.

More importantly, for payment purposes CMS is not interested in each HCC *per se*, it is interested in the sum of the HCCs along with demographic characteristics used to create the risk score. This means that positive and negative errors will generally offset, mitigating the impact of any bias.

Under this payment framework, while increasing the accuracy of the model may be appropriate, adjusting for underpayments while not adjusting for overpayments is inappropriate. Consider an example where HCCs have an average payment error of zero (i.e., unbiased, but a standard deviation of ten percent). It should be noted that plans are not paid for single HCCs. They are paid in the context of a portfolio of HCCs both within and across enrollees. In this circumstance, there will be some HCCs that have an underpayment of ten percent. However, because enrollees have a number of HCCs and plans have a number of enrollees, the error dramatically reduces by diversification.

Table 1. Reduction in Error by Diversification

Number of HCCs	Mean	Error
1	0	0.100
2	0	0.071
3	0	0.041
4	0	0.020
5	0	0.009
6	0	0.004
7	0	0.001

We see that multiple HCC submissions dramatically reduce the error rate. In addition, if we attempted to adjust on only correct underpayments, we would introduce an overpayment bias. The regression estimates, for our purposes of estimating the relationship between Medicare payments and risk scores, must follow:

$$\overline{Exp} = \alpha\overline{D} + \beta\overline{HCC}$$

Where $\overline{Exp}, \overline{D}$ and $\overline{HCC}$ represent the model averages. The above equation implies that:

$$\alpha\overline{D} = \overline{Exp} - \beta\overline{HCC}$$

Thus, if the coefficients on the HCCs were biased downward, the demographic coefficients would be forced upward to maintain the expenditure average. Since risk scores sum the demographic factors as well as the HCCs, this would have a moderating effect on any bias. As a consequence, it is unlikely that a bias would occur and because of

the purpose for which the model is used, econometric corrections might not be the appropriate remedy.[2]

II. Calibration Error Correction Limited to Recoveries is Economically Problematic

If there was a payment bias due to audit miscalibration, an adjustment targeted only to plans in a RADV audit would be arbitrary. An adjustment to original payments would be more appropriate.

Why Direct (Mean-Based) Application of the FFSA is Inappropriate

Assume that an enrollee's risk score is r_i^o. The enrollee is paid at the county rate C_i. The total original payment is $r_i^o C_i$. Upon audit, the risk score is recalculated to r_i^c. The enrollee's payment error (PE_i) is $(r_i^o - r_i^c)C_i$. Positive values indicate that the plan was overpaid and vice versa for negative values.

The calibration error (i.e., the error from using unaudited FFS data to develop the risk scores) presumably causes risk scores to have error due to the model being calibrated on unaudited FFS data. If Δ is the percent error in risk score due to calibration error, then the payment error, accounting for calibration error, is:

$$\left[\left(r_i^o(1-\Delta)\right) - \left(r_i^c(1-\Delta)\right)\right]C_i = \left[(r_i^o - r_i^c)C_i - \Delta(r_i^o - r_i^c)C_i\right] = PE_i - \Delta PE_i$$

Thus, the corrected payment error is:

$$PE_i^{**} = (1-\Delta)PE_i$$

In order for the direct adjustment to reduce the payment error, the Δ must be positive. But if Δ is positive, plans were overpaid in the first place. This leads to overpaid plans getting reduced recoveries. For example, if a representative MA population had an R^O equal to 1.045 and an R^c of 1.030, Δ would equal 1.5% ((1.045/1.030) – 1.0). Thus, the PE^* would need to be reduced by 0.985 (1-0.015):

$$PE^{**} = .985 * PE^*$$

A counterintuitive result arises when applying the adjustment in this scenario: the adjustment reduces the payment error amount. That is, MA plans were overpaid in the first place, and then they are being provided an offset to the payment error amount, which can be viewed as providing the plan an additional improper payment.

Now, consider the circumstance where plans are underpaid because the original unaudited model calibrated risk scores were too low. In this circumstance, Δ is negative

[2] For example, the FFS and MA populations are different. A bias may appear because of differences in the manner in which HCCs are utilized in the FFS and MA populations. This would have nothing to do with measurement error bias.

and the adjustment will increase the payment error. This is problematic as the original plan payments are not increased to compensate. Applying the adjustment only to the payment error would damage plans, as they would have been paid less than their enrollees' risk scores warrant and now the Δ adjustment (i.e., FFSA) would actually take back more money as well. This is again a counterintuitive result.[3]

In summary, if a plan's original payments represent consistent underpayments, the correct economic solution is to pay plans for their underpayments. Thus, eliminating a need to adjust recoveries. Moreover, if plans were consistently underpaid, and the attempt was made to collect the true economic value of the recoveries, those plans would have to return an amount over and above their original payment from CMS. In this case, plans would be returning payments to CMS that they never received which is counterintuitive.

To discount underpaid plans based directly on the audit miscalibration bias is contrary to economic reasoning. While its effect would moderate the impact of the original underpayment, the impacts would be economically arbitrary.

III. Empirical Analysis

This section describes the empirical analysis.

A. Input Data

This analysis requires three source data inputs, (a) FFS calibration data, (b) MA enrollee data and (c) the probability of HCC discrepancy for a RADV-like audit.

1. FFS data

The FFS data used in this analysis was the 2004-2005 file used for model calibration. There are 1,441,247 records in this data set.

2. MA sample

The MA sample has two million records that were sampled from the 2011 overpayment run. Approximately, one-half of the records are randomly sampled from the RADV eligible population and the other half are sampled from the non-RADV eligible population (excluding new, ESRD, and hospice enrollees).

3. Comprehensive Error Rate Testing (CERT)

[3] Say that the government purchased a $4 widget for $3. The government later wanted to get its money back because the widget was bad. A widget adjuster might say that there is a $1 bias underpayment. The correct economic value is (original price+bias = $3+$1) = $4. For the government to use the widget adjuster to correct the widget price to true economic value and demand $4 when they originally paid $3 is not equitable.

A sample of FFS claims and associated medical records were collected from the (CERT) audit. The CERT audit samples FFS claims for the purpose of reporting the error rate in FFS procedure code reporting. The sample consisted of a subset of claims with CY 2008 dates of service and the associated medical records.

B. Methodology

1. Probability of FFS Discrepancy

FFS Claim Level Discrepancy rates

Using the associated medical records, CPI performed a RADV-like review on the CERT data. We filtered the CERT data to create a subset of 8,630 unique claims to target for a RADV-like medical record review. The filtering (inclusion) criteria included:

- Outpatient claims paid through Medicare Part B. While CERT also samples Part A claims, they were not included in the FFS08 sample
- Claims that originated from an acceptable risk adjustment provider type
- Claims that contained ICD-9-CM diagnosis code(s) that mapped to one or more CMS-HCC

We will refer to the claims targeted for the audit as FFS08. CERT provided CPI with medical records for every claim in the sample. Using RADV coding guidelines, we performed a medical record review to develop discrepancy rates for the HCCs mapped from diagnoses on the claims. Table 2a summarizes the findings of the medical record review on FFS08.

Table 2a: FFS08 CERT Sample RADV HCC Discrepancy Counts at Claim Level

HCC	HCC Label	Sampled	Confirmed	Discrepant	%Discrepant
ALL	ALL	8,763	5,790	2,973	33.90%
HCC1	HIV/AIDS	22	16	6	27.30%
HCC2	Septicemia/Shock	51	30	21	41.20%
HCC5	Opportunistic Infections	12	3	9	75.00%
HCC7	Metastatic Cancer and Acute Leukemia	74	53	21	28.40%
HCC8	Lung, Upper Digestive Tract, and Other Severe Cancers	246	149	97	39.40%
HCC9	Lymphatic, Head and Neck, Brain and Other Major Cancers	275	170	105	38.20%
HCC10	Breast, Prostate, Colorectal and Other Cancers and Tumors	790	444	346	43.80%
HCC15	Diabetes with Renal or Peripheral Circulatory Manifestation	104	87	17	16.30%
HCC16	Diabetes with Neurologic or Other Specified Manifestation	123	98	25	20.30%
HCC17	Diabetes with Acute Complications	5	5	0	0.00%
HCC18	Diabetes with Ophthalmologic or Unspecified Manifestation	64	57	7	10.90%
HCC19	Diabetes without Complication	1,122	895	227	20.20%
HCC21	Protein-Calorie Malnutrition	11	5	6	54.50%

HCC	HCC Label	Sampled	Confirmed	Discrepant	%Discrepant
HCC25	End-Stage Liver Disease	4	3	1	25.00%
HCC26	Cirrhosis of Liver	19	14	5	26.30%
HCC27	Chronic Hepatitis	11	6	5	45.50%
HCC31	Intestinal Obstruction/Perforation	56	41	15	26.80%
HCC32	Pancreatic Disease	36	30	6	16.70%
HCC33	Inflammatory Bowel Disease	33	25	8	24.20%
HCC37	Bone/Joint/Muscle Infections/Necrosis	45	32	13	28.90%
HCC38	Rheumatoid Arthritis and Inflammatory Connective Tissue Disease	219	161	58	26.50%
HCC44	Severe Hematological Disorders	76	40	36	47.40%
HCC45	Disorders of Immunity	32	4	28	87.50%
HCC51	Drug/Alcohol Psychosis	4	2	2	50.00%
HCC52	Drug/Alcohol Dependence	16	9	7	43.80%
HCC54	Schizophrenia	197	111	86	43.70%
HCC55	Major Depressive, Bipolar, and Paranoid Disorders	448	216	232	51.80%
HCC67	Quadriplegia, Other Extensive Paralysis	4	0	4	100.00%
HCC68	Paraplegia	7	6	1	14.30%
HCC69	Spinal Cord Disorders/Injuries	8	7	1	12.50%
HCC70	Muscular Dystrophy	1	1	0	0.00%
HCC71	Polyneuropathy	81	43	38	46.90%
HCC72	Multiple Sclerosis	33	22	11	33.30%
HCC73	Parkinson's and Huntington's Diseases	69	51	18	26.10%
HCC74	Seizure Disorders and Convulsions	112	85	27	24.10%
HCC75	Coma, Brain Compression/Anoxic Damage	4	1	3	75.00%
HCC77	Respirator Dependence/Tracheostomy Status	0			
HCC78	Respiratory Arrest	7	5	2	28.60%
HCC79	Cardio-Respiratory Failure and Shock	234	170	64	27.40%
HCC80	Congestive Heart Failure	519	363	156	30.10%
HCC81	Acute Myocardial Infarction	41	29	12	29.30%
HCC82	Unstable Angina and Other Acute Ischemic Heart Disease	80	54	26	32.50%
HCC83	Angina Pectoris/Old Myocardial Infarction	103	54	49	47.60%
HCC92	Specified Heart Arrhythmias	900	523	377	41.90%
HCC95	Cerebral Hemorrhage	24	14	10	41.70%
HCC96	Ischemic or Unspecified Stroke	139	56	83	59.70%
HCC100	Hemiplegia/Hemiparesis	22	15	7	31.80%
HCC101	Cerebral Palsy and Other Paralytic Syndromes	3	1	2	66.70%
HCC104	Vascular Disease with Complications	105	70	35	33.30%
HCC105	Vascular Disease	444	297	147	33.10%

HCC	HCC Label	Sampled	Confirmed	Discrepant	%Discrepant
HCC107	Cystic Fibrosis	1	0	1	100.00%
HCC108	Chronic Obstructive Pulmonary Disease	484	388	96	19.80%
HCC111	Aspiration and Specified Bacterial Pneumonias	29	18	11	37.90%
HCC112	Pneumococcal Pneumonia, Emphysema, Lung Abscess	12	6	6	50.00%
HCC119	Proliferative Diabetic Retinopathy and Vitreous Hemorrhage	24	19	5	20.80%
HCC130	Dialysis Status	18	15	3	16.70%
HCC131	Renal Failure	604	384	220	36.40%
HCC132	Nephritis	8	5	3	37.50%
HCC148	Decubitus Ulcer of Skin	65	49	16	24.60%
HCC149	Chronic Ulcer of Skin, Except Decubitus	185	148	37	20.00%
HCC150	Extensive Third-Degree Burns	1	1	0	0.00%
HCC154	Severe Head Injury	2	1	1	50.00%
HCC155	Major Head Injury	16	6	10	62.50%
HCC157	Vertebral Fractures without Spinal Cord Injury	41	32	9	22.00%
HCC158	Hip Fracture/Dislocation	135	77	58	43.00%
HCC161	Traumatic Amputation	2	2	0	0.00%
HCC164	Major Complications of Medical Care and Trauma	73	43	30	41.10%
HCC174	Major Organ Transplant Status	7	5	2	28.60%
HCC176	Artificial Openings for Feeding or Elimination	20	18	2	10.00%
HCC177	Amputation Status, Lower Limb/Amputation Complications	1	0	1	100.00%

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. [4]

In this context, variation in HCC discrepancy rates could vary for two reasons. First, there are attributes intrinsic to a specific HCC that drive measurement error higher or lower. For example, the underlying diseases may be more difficult to evaluate and therefore the discrepancy rate or measurement error would be higher. In addition to measurement error, insufficient sample size can also contribute to the variation in the discrepancy rate. The purpose of this analysis is to determine the effects of the measurement error on the model calibration. Variation due to insufficient sample size confounds the analysis. We mitigated the impact of small sample size on the analysis as follows. While sample size is an issue in estimating HCC specific error rates, we do have sufficient data to calculate a reliable estimate of an overall HCC expected discrepancy

[4] The minimum number needed to determine a difference from zero at 90 percent confidence with a 0.15 margin of error.

rate. This is calculated by dividing the total number of HCCs sampled by the number of HCCs found to be discrepant. The overall discrepancy rate contains the most information as it is derived from all HCCs. Thus, in the absence of better information, it is the best measure of the probability of finding an HCC discrepant due to a RADV audit.
Using the overall discrepancy rate as our baseline estimate of HCC discrepancy, we can statistically determine, in a context that controls for sample size effects, if the HCC specific rates are significantly different from the overall rate. We apply a statistical test to each HCC level discrepancy rate to determine if it is statistically different at a 95 percent level of confidence from the overall rate. This approach confirms that differences from the overall discrepancy rate are likely due to intrinsic qualities of the HCC rather than small sample size.

When the individual rate is found to be statistically greater than the baseline rate, we assign the HCC to a high category. Conversely, we assign the HCC to a low category when the individual rate is statistically lower than the baseline rate. Rates found to be neither higher nor lower are assigned to the norm category. The discrepancy rates used in the analysis would then be assigned as follows. HCCs in the high and low category would use the mean of the respective category. HCCs in the norm category would use the baseline rate. Following this approach, the discrepancy rates for the baseline, high and low were 0.34, 0.46, and 0.21 respectively. Table 2b, reports the adjusted claim level error rate for each HCC.

FFS Beneficiary Level Discrepancy rates

Each enrollee HCC potentially has multiple claims with independent supportive medical records. Consequently, more medical claims imply the existence of more medical records to confirm an HCC. Clearly, the effective probability of discrepancy depends on the number of claims.

Based on the above, we need to establish the probability that an HCC will be in error for an FFS enrollee given the expected number of associated claims. We will refer to this as the beneficiary level discrepancy rate. As previously described, we established base probabilities that each HCC j will be in error for a given claim, p_j. In table 2b, for each HCC j, we also establish the average number of underlying claims, $\bar{z}$. Accordingly, the beneficiary level discrepancy rate is:

$$p_j^{\bar{z}} = [Claim\ Level\ Discrepancy\ Rate]^{[Average\ Claims]}$$

Table 2c shows the distribution of the beneficiary level discrepancy rates. The adjustment of the claim level discrepancy rate to a beneficiary level discrepancy rate results in a much smaller probability of an HCC being in error as a result of FFS diagnoses error in the FFS claims. As shown in the table, seventy-five percent of the beneficiary level discrepancy rates are 5% or below. The average rate is 3% and the median rate is 2%.

Table 2b: Statistical HCC Base Discrepancy Probabilities and Effective Error Rates

HCC	Cat	Average Number of Claims per HCC* ($\bar{Z}$)	Claim Level Discrepancy Rate**	Beneficiary Level Discrepancy Rate ($p_j^{\bar{z}}$)
1	N	11.4	33.8%	0.000%
2	N	4.1	33.8%	1.223%
5	H	2.7	46.2%	12.259%
7	N	7.0	33.8%	0.050%
8	N	13.4	33.8%	0.000%
9	N	11.0	33.8%	0.001%
10	H	7.0	46.2%	0.432%
15	L	3.3	20.9%	0.581%
16	L	3.3	20.9%	0.581%
17	N	2.6	33.8%	5.716%
18	L	2.3	20.9%	2.888%
19	L	6.2	20.9%	0.006%
21	N	2.5	33.8%	6.814%
25	N	3.7	33.8%	1.739%
26	N	5.2	33.8%	0.349%
27	N	3.2	33.8%	3.127%
31	N	3.5	33.8%	2.278%
32	L	3.7	20.9%	0.285%
33	N	3.7	33.8%	1.898%
37	N	4.7	33.8%	0.621%
38	L	4.9	20.9%	0.045%
44	H	6.1	46.2%	0.891%
45	H	3.8	46.2%	5.332%
51	N	2.2	33.8%	8.717%
52	N	3.6	33.8%	2.084%
54	H	9.0	46.2%	0.098%
55	H	6.1	46.2%	0.889%
67	N	4.5	33.8%	0.726%
68	N	4.1	33.8%	1.225%
69	N	2.4	33.8%	7.163%
70	N	3.3	33.8%	2.851%
71	H	2.6	46.2%	12.931%
72	N	7.3	33.8%	0.037%
73	N	5.2	33.8%	0.340%
74	L	4.7	20.9%	0.063%

HCC	Cat	Average Number of Claims per HCC* ($\bar{Z}$)	Claim Level Discrepancy Rate**	Beneficiary Level Discrepancy Rate ($p_j^{\bar{z}}$)
75	N	2.4	33.8%	7.669%
77	N	3.2	33.8%	3.100%
78	N	2.0	33.8%	11.297%
79	L	5.0	20.9%	0.040%
80	N	6.1	33.8%	0.138%
81	N	3.6	33.8%	2.057%
82	N	2.9	33.8%	4.205%
83	H	2.4	46.2%	15.435%
92	H	7.0	46.2%	0.446%
95	N	4.7	33.8%	0.630%
96	H	4.1	46.2%	4.186%
100	N	3.2	33.8%	3.188%
101	N	2.6	33.8%	5.672%
104	N	4.0	33.8%	1.241%
105	N	3.4	33.8%	2.389%
107	N	6.5	33.8%	0.084%
108	L	5.0	20.9%	0.041%
111	N	3.2	33.8%	3.248%
112	N	2.3	33.8%	7.966%
119	N	2.6	33.8%	5.710%
130	L	4.5	20.9%	0.090%
131	N	4.7	33.8%	0.593%
132	N	2.3	33.8%	8.185%
148	N	2.7	33.8%	5.642%
149	L	4.8	20.9%	0.051%
150	N	2.9	33.8%	4.372%
154	N	2.3	33.8%	7.897%
155	H	3.1	46.2%	8.908%
157	N	3.6	33.8%	2.114%
158	H	7.5	46.2%	0.295%
161	N	3.1	33.8%	3.593%
164	N	2.6	33.8%	5.886%
174	N	10.1	33.8%	0.002%
176	L	2.7	20.9%	1.359%
177	N	3.4	33.8%	2.495%

*The average claim per HCC was calculated using the underlying claims file for the 2004-2005 calibration data set.

**Rate as adjusted for sample size.

Table 2c: Summary Statistics for Effective Probabilities

MAX	15.435%
MIN	0.000%
MEDIAN	1.819%
AVERAGE	3.064%
1st QUARTILE	0.306%
3rd QUARTILE	5.092%

2. Audit Miscalibration Impact

Theoretical Impact to Risk Scores

A key assertion justifying the FFS adjuster has been the assumption that if the underlying FFS claims were subject to a RADV-like audit that resulted in one or more beneficiary HCC changing from 1 to 0, the HCC risk factors from the re-calibrated CMS-HCC model would increase. The argument would then follow that the increased risk factors would increase a plan's payment.

To understand why this is incorrect, we must recognize how the CMS-HCC model is used in payment. First, CMS derives the base capitation rates by county and largely independent of the CMS-HCC model.[5] This establishes the fundamental payment. The CMS-HCC model is then used to adjust those payments by predictive disease-cost load. For example, assume that in a particular county the capitated payment is $\$X$.[6] Plans in that county have the incentive to directly or indirectly select enrollees that will cost less than $\$X$ and exclude those likely to cost more than $\$X$ (adverse selection). Noticeably, plans that have a greater number of likely more expensive enrollees are at greater financial risk. Finally, this adverse selection will increase the cost to the Medicare Trust Fund. To combat this adverse selection and financial risk, CMS created a risk score (R) for each enrollee that will adjust base payments to pay more for likely more expensive enrollees and less for likely less expensive enrollees. So a plan receiving $\$X$ for each enrollee will instead receive $\$X*R$. Since an average cost enrollee will have a risk score of one (R=1) the plan will still receive $\$X$ ($\$X*R=\$X*1=\$X$) for an average enrollee. The CMS-HCC model bases the risk score (R) on the possession of demographic and chronic disease characteristics. Although fundamentally based on expenditures, the regression is adjusted such that the HCC and demographic factors will provide an average risk score of one on the calibrating FFS dataset. Thus, by definition of the CMS-HCC model, the average FFS enrollee will never have a risk score of greater or less than one.

On the other hand, the average MA enrollee risk score is not constrained to any particular value. If the MA population is generally unhealthy relative to the FFS population, the average risk score for an MA enrollee will be greater than one. If the MA population is

[5] The county rates are primarily derived from FFS data and the bidding process.
[6] To simplify the discussion, for illustrative purposes we refer to a single county rate. In the actual payment there is a specific rate for each plan per county.

generally healthier than the FFS population, the average risk score for an MA enrollee will be less than one.

Collectively, the calibration of the CMS-HCC model cannot, by definition, cause the risk factors to increase such that the average risk score of the calibrating FFS data is greater than one. Therefore, calibrating the CMS-HCC model on an audited dataset would simply change the risk factors relative to each other. It is conceptually possible that the MA population utilizes more of the risk factors that increase under the audited calibration methodology and thus increase MA payments. However, there is no theoretical reason for this to be the case. Nonetheless, we will examine the issue empirically.

Impact Measurement

The objective of the analysis is to compare the risk scores that come out of the original uncorrected data model to those that come out of the RADV-like corrected data model. The difference, when applied to MA enrollees, would be an estimate for the bias. To perform the analysis of the impact of the RADV-like audit, we proceed as follows. First, we estimate a proxy for the CMS-HCC model on the original uncorrected dataset and estimate the original factors. We then apply these original factors to the MA sample and estimate original risk scores for the uncorrected calibrating dataset.

Next, we carry out a series of steps to simulate the same MA enrollees' corrected risk scores that would have resulted from calibrating the model on a FFS dataset that was RADV-like corrected.

We apply these probabilities to the calibrating dataset in such a manner that each existing HCC of each enrollee is subject to a $p_j^{\bar{\bar{z}}}$ probability of being redefined as discrepant. We then estimate the CMS-HCC model on the simulated corrected data. In the next step, we take the new coefficients and apply them on the original FFS data set, normalizing a new set of relative factors to one. The new relative factors are then applied to the beneficiary profiles in the MA sample to calculate new risk scores based on coefficients unaffected by FFS diagnoses error in the calibration. All other things being equal, the difference between the error-free risk score and the original risk score is the impact of FFS diagnoses error in the calibration. The normalization process is illustrated as follows: A simplified risk adjustment model is

$$E = \alpha DEM + \beta HCC1 + \delta HCC2 + \epsilon$$

Here, E is the enrollee expenditures from period (t), DEM is a demographic characteristic (such as sex (male=1, female=1)), *HCC1* and *HCC2* are equal to one if the particular condition is present and zero if it is not present at time (*t-1*) and ϵ is a error term satisfying the assumptions of the Ordinary Least Squares Model.

Since E is dollar expenditures, the estimates coefficient of the model $\hat{\alpha}, \hat{\beta}, \hat{\delta}$ will also be dollar expenditures. $\hat{\alpha}$ is the incremental dollar cost of being female over male; $\hat{\beta}$ is the incremental dollar cost of having HCC1 and $\hat{\delta}$ is the incremental dollar cost of having

HCC2. We start with an enrollee profile of five FFS enrollees. The actual FFS expenditures for enrollees 1-5 are E1, E2, E3, E4 and E5.

$$\begin{bmatrix} E & DEM & HCC1 & HCC2 \\ E1 & 1 & 0 & 1 \\ E2 & 1 & 1 & 1 \\ E3 & 0 & 1 & 1 \\ E4 & 1 & 0 & 1 \\ E5 & 1 & 1 & 0 \end{bmatrix}$$

We estimate the model and obtain estimated coefficient of $\hat{\alpha}, \hat{\beta}, \hat{\delta}$ which are dollar estimates. To create the risk scores, we apply these estimates to the enrollee profiles derived from the uncorrected FFS data.

$$\hat{\boldsymbol{X}} = \begin{bmatrix} Bene & DEM & HCC1 & HCC2 \\ 1 & \hat{\alpha} & 0 & \hat{\delta} \\ 2 & \hat{\alpha} & \hat{\beta} & \hat{\delta} \\ 3 & 0 & \hat{\beta} & \hat{\delta} \\ 4 & \hat{\alpha} & 0 & \hat{\delta} \\ 5 & \hat{\alpha} & \hat{\beta} & 0 \end{bmatrix}$$

Then we sum the expected expenditure for each enrollee:

$Bene$	Sum	$Estimate$
1	$\hat{\alpha} + \hat{\delta}$	$= \hat{E}1$
2	$\hat{\alpha} + \hat{\beta} + \hat{\delta}$	$= \hat{E}2$
3	$\hat{\beta} + \hat{\delta}$	$= \hat{E}3$
4	$\hat{\alpha} + \hat{\delta}$	$= \hat{E}4$
5	$\hat{\alpha} + \hat{\beta}$	$= \hat{E}5$

To normalize into risk scores, we first obtain the average predicted expenditure of the enrollees:

$$\bar{E} = \frac{\hat{E}1 + \hat{E}2 + \hat{E}3 + \hat{E}4 + \hat{E}5}{5}$$

Next, we divide the expected expenditures for each enrollee by the average expenditure of all enrollees:

Bene	*Sum*	*Risk Score*
1	$\frac{\hat{E}1}{\bar{E}}$	$= R1$
2	$\frac{\hat{E}2}{\bar{E}}$	$= R2$
3	$\frac{\hat{E}3}{\bar{E}}$	$= R3$
4	$\frac{\hat{E}4}{\bar{E}}$	$= R4$
5	$\frac{\hat{E}5}{\bar{E}}$	$= R5$

The normalized risk scores $R1, R2, R3, R4, R5$ must, on average, equal one. It is important to note that this is true regardless of the enrollee profile used to calibrate the model and the level of "correctness" of the enrollee profile. It also follows that the estimates $\hat{\alpha}, \hat{\beta}, \hat{\delta}$ will always be such that the calibrating dataset's sum of actual expenditures will be equal to the expected expenditures. Consistent with the above, the risk factor for *DEM* is $\frac{\hat{\alpha}}{\bar{E}}$, for *HCC1* is $\frac{\hat{\beta}}{\bar{E}}$ and for *HCC2* is $\frac{\hat{\delta}}{\bar{E}}$. Finally, corrected factors are applied to the MA enrollee sample to establish the corrected MA risk scores. For each enrollee k in the calibrated dataset, we calculate the percentage difference in risk score from the original score:

$$d_k = (corrected\ risk\ score_k - original\ risk\ score_k)/original\ risk\ score_k$$

If there are N enrollees in the MA sample dataset, then the average of the percentage difference in risk score across these enrollees is an estimate of the bias,

$$\bar{d}_k = \frac{1}{N}\sum_{k=1}^{N} d_k$$

For the bias to be impactful against the MA plans, it should be a large, positive percentage.

3. Simulation and Empirical Results

In order to estimate variation in $\bar{d}_k$, we ran a simulation calculating 50 cases of the impact analysis. For each case, using a random number generator, we applied the beneficiary level discrepancy rates to get 50 independent corrected FFS data files. Each file was then used to re-calibrate the CMS-HCC model, re-normalize a new set of risk scores and estimate a new value of $\bar{d}_k$. This resulted in a distribution of 50 bias estimates, $\bar{d}_k^1, , \bar{d}_k^2, , \bar{d}_k^3, \cdots, , \bar{d}_k^{50}$.

This distribution of relative differences is reported below. The mean bias estimate of these 50 simulations was -0.08% and the mean bias estimate was between -.07%,-.09% at a 95% level of confidence.

An examination of Table 3 and Figure 1 makes it clear that the chance of any one of these simulations being greater than zero is very unlikely. While there appears to be a negative bias, it is so negligible that it should be considered to be zero. This shows that FFS diagnosis error in the FFS calibration data does not manifest a negative payment bias to MA plans.

Table 3. Distributional Estimate of Bias

Distributional Statistic	Bias Estimate
Mean Relative Difference	-0.08%
Median Relative Difference	-0.09%
Minimum Relative Difference	-0.23%
Maximum Relative Difference	0.01%
25th Percentile	-0.10%
7th Percentile	-0.05%

Figure 1 Sampling Distribution of the Relative Difference