Medicare & Medicaid Research Review
2014: Volume 4, Number 2

*A publication of the Centers for Medicare & Medicaid Services, Office of Information Products & Data Analytics*

# Measuring Coding Intensity in the Medicare Advantage Program

*Richard Kronick[1] and W. Pete Welch[2]*
[1]Department of Health and Human Services—Agency for Healthcare Research and Quality
[2]Department of Health and Human Services—Office of the Assistant Secretary for Planning and Evaluation

**Background:** In 2004, Medicare implemented a system of paying Medicare Advantage (MA) plans that gave them greater incentive than fee-for-service (FFS) providers to report diagnoses.

**Data:** Risk scores for all Medicare beneficiaries 2004–2013 and Medicare Current Beneficiary Survey (MCBS) data, 2006–2011.

**Measures:** Change in average risk score for all enrollees and for stayers (beneficiaries who were in either FFS or MA for two consecutive years). Prevalence rates by Hierarchical Condition Category (HCC).

**Results:** Each year the average MA risk score increased faster than the average FFS score. Using the risk adjustment model in place in 2004, the average MA score as a ratio of the average FFS score would have increased from 90% in 2004 to 109% in 2013. Using the model partially implemented in 2014, the ratio would have increased from 88% to 102%. The increase in relative MA scores appears to largely reflect changes in diagnostic coding, not real increases in the morbidity of MA enrollees. In survey-based data for 2006–2011, the MA-FFS ratio of risk scores remained roughly constant at 96%. Intensity of coding varies widely by contract, with some contracts coding very similarly to FFS and others coding much more intensely than the MA average. Underpinning this relative growth in scores is particularly rapid relative growth in a subset of HCCs.

**Discussion:** Medicare has taken significant steps to mitigate the effects of coding intensity in MA, including implementing a 3.4% coding intensity adjustment in 2010 and revising the risk adjustment model in 2013 and 2014. Given the continuous relative increase in the average MA risk score, further policy changes will likely be necessary.

**Keywords:** Medicare, payment systems, FFS, Capitation, risk adjusted payments, Managed Care Organizations

ISSN: 2159-0354

doi: http://dx.doi.org/10.5600/mmrr.004.02.a06

Medicare & Medicaid Research Review
2014: Volume 4, Number 2

**Mission Statement**

*Medicare & Medicaid Research Review* is a peer-reviewed, online journal reporting data and research that informs current and future directions of the Medicare, Medicaid, and Children's Health Insurance programs. The journal seeks to examine and evaluate health care coverage, quality and access to care for beneficiaries, and payment for health services.

http://www.cms.gov/MMRR/



Published by the
Centers for Medicare & Medicaid Services.



# Introduction

## General

Enrollment in the Medicare Advantage Program has increased dramatically, growing from 9.3 million beneficiaries in March 2008 to 15.4 million in March 2014, an increase of 66% in just six years.

Concerns about overpayment as a result of favorable risk selection have confronted the Medicare program throughout the history of Medicare contracting with health maintenance organizations and other private plans. In the late 1980s, Medicare paid health plans using a system that adjusted for demographic factors such as age and gender, but plan enrollees were healthier than fee-for-service beneficiaries with the same demographic characteristics, and, as a result, health plans were estimated to be overpaid by approximately 11% (Brown, Bergeron, Clement, Hill, & Retchin, 1993).[1]

In order to reward health plans for attracting sicker-than-average enrollees, and to discourage plans from constructing business models designed to avoid risk, Medicare and other payers have increasingly turned to diagnosis-based risk-adjusted payment systems in which health plans are paid more for enrollees expected to need more care. While mitigating the incentive to enroll only healthy people, diagnosis-based risk adjustment creates another set of incentives: to find and report as many diagnoses as possible. Risk-adjusted payment using plan-reported diagnostic information creates a dilemma—there are strong policy reasons to pay plans more if they enroll sicker people,

[1] Systems in which payment is based, at least in part, on patient diagnoses create an incentive for providers or plans to report all appropriate diagnoses. Most notably, after the advent of the inpatient prospective payment system, which used diagnosis related groups (DRGs), hospital case-mix increased.

but measuring morbidity using plan-reported diagnostic data provides strong incentives for plans to increase the reported morbidity of enrollees (Government Accountability Office, 2012).

## Medicare Risk Adjustment

The first Medicare risk program was established under the Tax Equity and Fiscal Responsibility Act (TEFRA) of 1982. From 1982 to 1999, this program paid health plans (known as TEFRA health maintenance organizations [HMOs]) a monthly amount per enrollee based on 95 percent of the estimated costs of treating an average beneficiary in the traditional fee-for-service (FFS) program. This amount was adjusted for demographic factors such as age and gender (Hellinger & Wong, 2000; Lichtenstein, Thomas, Adams-Watson, Lepkowski, & Simone, 1991).

The Balanced Budget Act of 1997 made significant changes under the Medicare+Choice (M+C) program, including requiring the development of a new risk adjustment system that incorporated morbidity information. Medicare began using diagnoses from inpatient hospitalizations in 2000 to adjust payments to health plans. The Benefits Improvement Protection Act of 2000 required the use of ambulatory diagnoses in Medicare risk-adjustment, and Medicare implemented the CMS-HCC (Centers for Medicare & Medicaid Services-Hierarchical Condition Categories) risk adjustment model in 2004 and fully phased it in by 2007. The CMS-HCC incorporated both inpatient and ambulatory diagnoses that were categorized into 70 HCCs that were determined to be predictive of costs (Pope, *et al*., 2004). In 2003 the Medicare Modernization Act (MMA) renamed the program "Medicare Advantage (MA)."

The MA payment system uses diagnostic information to assign a risk score to each beneficiary, where the average beneficiary in fee-for-service has a risk score of 1.0. MA plans are paid the product of their bid multiplied by the enrollee's risk score—that is, if an MA plan bids $1,000/month for an enrollee with a risk score of 1.0, and then enrolls a beneficiary with a risk score of 1.2, the plan gets paid $1,200/month for that enrollee (1.2 * $1,000/month).

This payment system creates incentives for MA plans to find and report as many diagnoses as can be supported by the medical record. For office-based services, physicians billing FFS are paid based on the procedures performed—for example, one amount for an intermediate office visit, a different amount for a skin biopsy—but are not paid based on the number of diagnoses that are reported (provided there is at least one that supports the need for the service provided). If, for example, a FFS patient with quadriplegia and a urinary tract infection (UTI) has an office visit for treatment of the UTI, the payment to the physician will be no more if both the UTI and quadriplegia are reported on the claim than if only the UTI is reported. Coding guidelines specify that the quadriplegia can legitimately be coded if it contributes to the complexity of care, but unlike the situation for MA plans, in FFS there is no incentive to report more than one diagnosis. In contrast, MA plans have a strong incentive to do so. In addition to the incentives to report more completely, the method of collecting diagnostic information also provides MA plans additional opportunities to increase risk scores. FFS diagnoses are drawn only from health care claims submitted for payment. MA plans may also review medical records and can report all diagnoses that are supported in the record, including those that were not reported by physicians on any health care claim or encounter record. MA plans can also employ nurses to visit enrollees in their homes to conduct health assessments and report diagnoses that are found.

As noted, payment to MA plans is calibrated based on coding patterns in FFS. If, for example, Jane Doe would have a risk score of 1.0 if she were in FFS, the implicit assumption in the design of the MA payment system is that she would have the same risk score of 1.0 if she were enrolled in MA. Alternatively, if the same Jane Doe had a risk score of 1.1 if enrolled in MA, then her plan would be overpaid. As we use the phrase, 'coding intensity' is the difference between the scores that a group of beneficiaries would have if enrolled in MA and their scores in FFS.

If we find that MA plans code with high intensity, that does not provide information about whether MA plans are 'overcoding' or whether FFS providers are 'undercoding.' Similarly, one cannot infer whether MA coding is any more or less accurate than FFS. However, for the purposes of creating an equitable MA payment system, the relevant question is whether MA risk scores are systematically different than those risk scores would be in FFS, and not which set of scores is more accurate. Suppose, for example, that average expenditure per FFS beneficiary is $10,000/year, and that the average risk score for all FFS beneficiaries is 1.0. The goal of the payment system is to assure that—if all of those beneficiaries were to enroll in MA—the average payment to MA plans would be approximately $10,000 per beneficiary. If, however, FFS systematically underreports diagnoses and MA reports them more completely, then the average risk score in MA would be greater than 1.0, and MA plans would be paid more than $10,000 per beneficiary. The goal of the risk adjustment system is to assure that MA plans that enroll sicker-than-average beneficiaries are paid appropriately, but not to increase payment for an average beneficiary.

As a result of concern about the effects of incentives to increase coding intensity on MA plan payment levels, the Deficit Reduction Act of 2005 directed CMS to measure and adjust for coding intensity (Department of Health and Human Services, 2007), and in the 2010 payment year CMS adjusted risk scores by 3.41% to reflect anticipated differences between MA and FFS coding (Department of Health and Human Services, 2009). The Government Accountability Office (2012) estimated that in 2010 MA beneficiary risk scores were at least 4.8% and perhaps as much as 7.1% higher than they likely would have been if the same beneficiaries had been continuously enrolled in FFS. The Affordable Care Act directs CMS to increase the coding intensity adjustment to at least 4.71% in 2014, and further increase it to at least 5.71% by 2018 (PPACA & HCERA, 2010). The American Taxpayer's Relief Act of 2012 further increases the minimum coding intensity adjustment to 4.91% in 2014 and 5.91% in 2018.

The risk adjustment model implemented in 2004 was recalibrated in 2007, 2009, and 2013.[2] These changes slowed the growth in measured MA risk scores somewhat. Further, as described in more detail below, since 2014, CMS has adopted substantial changes to the model, particularly relating to several diagnoses that have been subject to coding intensity efforts by MA plans (Department of Health and Human Services, 2013). In 2014 the risk score is a blend, weighting the risk score calculated using the 2013 model by 25% and the risk score calculated using the 2014 model by 75%.

Concerns about coding intensity in MA plans would be minor if coding in FFS were relatively complete, because in that case there would be little opportunity for MA plans to legitimately increase risk scores through efforts at increasing diagnostic

[2] HCCs are constrained for a variety of reasons, including methodological and policy reasons. Starting in 2013, the coefficients for the four most severe diabetes HCCs were constrained to the same amount, eliminating the opportunity to increase revenue by increasing the reported severity of diabetes within these four HCCs.

reporting. However, FFS coding is known to be both incomplete and variable. Incomplete coding is evidenced by lack of persistence in coding of chronic conditions. For instance, among Medicare beneficiaries diagnosed with quadriplegia in one year, only 61% had a diagnosis of quadriplegia reported in the subsequent year (Medicare Payment Advisory Commission, 1998). Coding intensity also varies geographically; in the Hospital Referral Regions (HRRs) with the most intense practice patterns, the probability that a beneficiary is diagnosed with three or more chronic conditions is double the probability in low-intensity HRRs, with no evidence of differences in underlying prevalence (Welch, Sharp, Gottlieb, Skinner, & Wennberg, 2011).

Incomplete and variable coding provide ample opportunities for MA plans to increase risk scores of beneficiaries through coding intensity efforts, and a number of vendors actively market services that help plans to do so (Gorman Health Group, 2013; Leprechaun, 2013), often advertising high returns on investment (ROIs) for their services.

This paper uses recent data as well as improved methods to estimate the effects of coding intensity on MA risk scores, with particular attention to variation over time in coding intensity, variation in coding intensity across plans, and variation in the diagnoses most subject to coding intensity efforts. Before describing our methods, we'll review what is known about risk selection in the MA program.

### *Evidence on Risk Selection*

Concerns about overpayment as a result of favorable risk selection have confronted the Medicare program throughout its history of contracting with HMOs and other private plans. For example, Brown *et al*. (1993) found that health plans were overpaid by approximately 11% under the Medicare risk program during the late 1980s because plans experienced favorable selection; on average, risk plan enrollees were healthier than fee-for-service beneficiaries with the same demographic characteristics, with lower prior reimbursements and fewer indicators of chronic health problems (Brown *et al*., 1993). Several studies confirmed that the favorable risk selection for early HMOs continued during the early 1990s (Mello, Stearns, Norton, & Ricketts, 2003; Call, Dowd, Feldman, & Maciejewski, 1999; Physician Payment Review Commission, 1996; Riley, Chiang, Ingber, & Tudor, 1996). For example, Mello *et al*. (2003) found that Medicare beneficiaries with a history of cancer or stroke were significantly less likely to be enrolled in HMOs.

Recent analyses suggest that the various policy changes that have been implemented over the past decade—such as the implementation of the CMS-HCC model, the addition of an open enrollment period, and additional refinements of the risk adjustment system—have reduced the degree of selection bias between Medicare Advantage (MA) and the traditional FFS program. For example, Newhouse, Price, Huang, McWilliams, & Hsu (2012) recently found that differences in risk scores for beneficiaries switching from the traditional FFS program to MA and beneficiaries remaining in the traditional program narrowed by a factor of 3 between 2004 and 2008 (with the difference between risk scores for beneficiaries switching into MA and those staying in the traditional FFS program changing from -0.113 to -0.037); and differences in adjusted mortality rates narrowed by a factor of 2 between 1988 and 2008 (with mortality among beneficiaries in MA as a percentage of mortality among beneficiaries in the traditional FFS program increasing from 85 percent to 93 percent) and were almost equal between persons enrolled in MA for five or more years and traditional FFS enrollees (Newhouse *et al*., 2012). Similarly, McWilliams, Hsu, and Newhouse (2012) found reductions in differences between MA and FFS for self-reported health and health care use between 2001 and 2003 and between 2006 and 2007. However, there continues to be evidence of selection bias in disenrollment from MA plans (Riley, 2012; Medicare Payment Advisory Commission, 2012; Morrisey, Kilgore, Becker, Smith, & Delzell, 2012).

## Methods

Ideally, we would measure coding intensity by running an experiment with a large group of beneficiaries, in which one-half of the group was randomly assigned to MA, the other half to FFS, and we observed the risk scores of each group. The difference between the average score in MA and FFS would be a good estimate of coding intensity. In the absence of performing this experiment, we use a variety of inferential approaches. As shown below, average MA risk scores increased much more rapidly than average FFS scores from 2004 to 2013. There are three potential explanations for the relative increase in MA scores. First, the composition of MA enrollment might have changed; for example, MA enrollees in 2013 might be older, relative to FFS, than MA enrollees in 2004, and, as a result have higher risk scores. Second, even if there were no change in the composition of MA enrollees, MA enrollees might have gotten sicker more quickly than FFS beneficiaries. Third, it is possible that coding intensity increased in MA. Below, we assess the plausibility of each of these three potential explanations concluding that greater coding intensity is by far the most likely explanation.

We evaluate whether changes in MA scores relative to FFS scores are due to coding intensity efforts or to changes in the composition of enrollees by analyzing the contribution of changes in risk scores for four types of beneficiaries: stayers, leavers, joiners, and switchers—where stayers are beneficiaries in either MA or FFS for two consecutive years; leavers are beneficiaries, primarily decedents, who were in one sector in the first year, but not Medicare eligible in the second year; joiners are beneficiaries, primarily those turning 65, who were not Medicare eligible in the first year; and switchers are beneficiaries who move from FFS to MA (or vice-versa) between two consecutive years. To the extent that the contribution of leavers, joiners, or switchers differed between MA and FFS, then we will have evidence that part of the differential growth in risk scores between MA and FFS is a result of differences arising from enrollment decisions or mortality of beneficiaries, or "caseload dynamics" for shorthand. However, if risk scores increase more rapidly for MA stayers than for similar FFS stayers, then we will have evidence that coding intensity accounts for part of the more rapid growth in MA scores.

Our analysis assumes that in the absence of coding intensity efforts, the risk score for a given MA beneficiary would increase, on average, at the same rate as the risk score for an otherwise similar FFS beneficiary from one year to the next. This appears to be a reasonable assumption, and would only be incorrect if MA beneficiaries were actually getting sicker more quickly than FFS beneficiaries. We analyze changes in the relative mortality rate of MA and FFS enrollees as an independent method of assessing the relative morbidity of beneficiaries in the two sectors. If MA enrollees are, in fact, getting sicker more quickly than FFS beneficiaries, we would expect to see MA mortality rates increase relative to FFS mortality.

## Data

For each year from 2004 to 2013, we used three Medicare administrative files with beneficiary level data: the Enrollment Database (EDB); the

Common Medicare Environment, which contains the MA contract number; and the Risk Adjustment Processing System, which contains the risk score. For each year from 2004 to 2013, we computed the risk score using each HCC's coefficient (i.e., weight) that was used to pay plans in 2004–2006. Analyses of coding intensity based on risk scores have not appeared in the peer-reviewed literature.

We excluded beneficiaries not enrolled in both Parts A and B, because diagnostic information would be incomplete. Beneficiaries enrolled in cost contracts, hospice, or the ESRD program were also excluded, because of differences in payment methodology. Risk scores were computed prior to any normalization and MA-wide adjustment for differences in coding.

## Decomposition of Risk Score Growth

To decompose the changes in risk scores, we analyze beneficiaries based on their enrollment as of July 1 of each of two consequent years (called a "cohort"). We analyze nine cohorts, from the first cohort under the new adjustment system in 2004–2005 through the most recent one in 2012–2013, estimating the contribution of each type of beneficiary—stayers, joiners, leavers, and switchers—to the overall change in risk score. The decomposition method is described in the Supplemental Appendix.

## Contract-Level Analyses

To assess the extent to which coding intensity differs across contracts, we calculate the rate of increase in risk scores for stayers for contracts that continuously participated in Medicare from 2004 through 2011, had at least 10,000 enrollees in both 2004 and 2011, and whose ratio of 2011 enrollment to 2004 enrollment was between 25% and 300%. The subset of contracts account for approximately 85% of enrollment in 2004 and 42% in 2011. For each contract and cohort, we define the excess increase in risk score for stayers as the difference between the increase in risk scores for stayers in the contract and the increase in risk scores for FFS beneficiaries, controlling for age and decedent status (details are in the Supplemental Appendix).[3] The cumulative excess increase in risk score is the sum of this amount over the seven cohorts from 2004 to 2005 through 2010 to 2011. Contracts are divided into deciles (weighted by 2011 enrollment) based on the cumulative excess increase in scores for stayers.

The growth in relative prevalence of each HCC was calculated separately for the MA and FFS sector as well as by decile of coding intensity among the continuously participating contracts.

## Analysis of Medicare Current Beneficiary Survey (MCBS) Data

We analyze the MCBS to provide an assessment of the relative risk of MA and FFS beneficiaries that is independent of the data reported by MA plans or FFS providers. As described more fully in the Supplemental Appendix, in the first stage of the analysis, we use MCBS data from 2004 to 2011 to estimate a regression model to predict the sum of Part A plus Part B expenditures. The regression is estimated on FFS enrollees. Independent variables are demographic characteristics (age, gender, Medicaid status, and institutional status), self-reported health status (excellent, very good, good, fair, or poor), and self-report of whether the respondent has ever been diagnosed with hypertension, heart attack, heart failure, stroke, cancer, or diabetes, and dummy variables for each year. In the second stage, we use the regression coefficients to predict expenditures for each FFS

[3] The contract-level analysis implicitly assumes that the increase in FFS risk scores for a beneficiary of a given age and decedent status is uniform geographically.

and MA respondent in the sample, and compare the average predictions for MA and FFS.

## Results

Mean risk scores in MA increased faster than in FFS from 2004 to 2013—by 0.305 in MA, compared to 0.106 in FFS (Exhibit 1). As measured using the risk model in effect in 2004, MA scores grew from 90.2% of FFS scores in 2004 to 109.1% in 2013, an increase of 18.9 percentage points.[4]

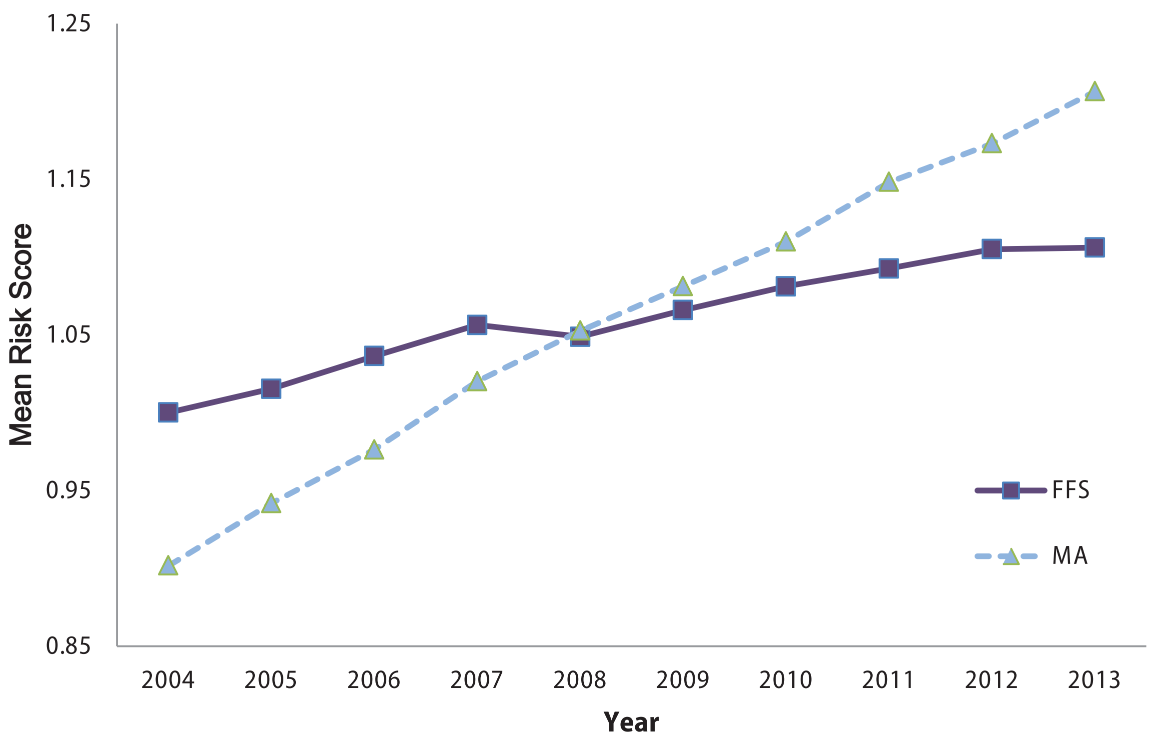
### Decomposition of Risk Score Growth

Risk scores among MA stayers increased more quickly than risk scores among FFS stayers in every cohort (Exhibit 2). Across the two-year cohorts, the growth in risk scores for FFS stayers averaged 0.088. In contrast, in MA the growth in mean risk scores for stayers was 0.106 in 2004–2006, 0.119 in 2006–2010, and 0.132 in 2010–2013. The difference between the rate of growth in risk scores for MA and FFS stayers is substantial, with average MA scores increasing one-third more rapidly than FFS scores.

As was shown in Exhibit 1, the average MA risk score increased 0.025 more rapidly than the average FFS score from 2004 to 2005. Differential changes between MA and FFS in caseload dynamics account for approximately one-half of this 0.025 difference (Exhibit 3). As was shown in Exhibit 2, risk scores for FFS stayers increase by approximately 0.088 per year. However, as was shown in Exhibit 1, the

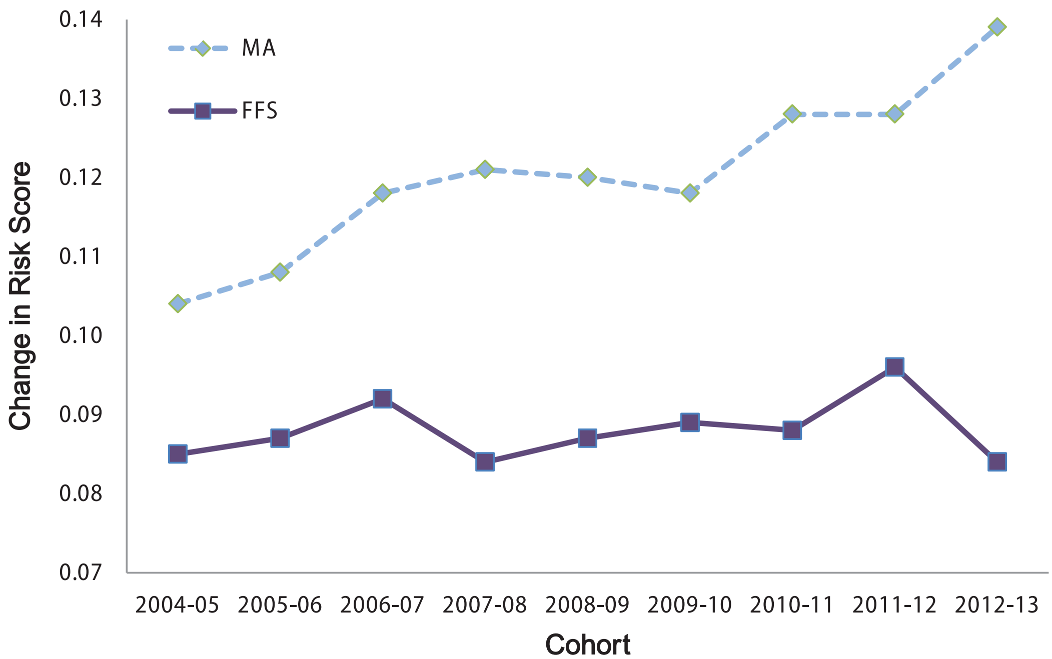
[4] As noted, the payment rules, however, did change in a variety of ways.

**Exhibit 1. Mean Risk Scores, MA vs. FFS, 2004–2013 (2004 Model)**



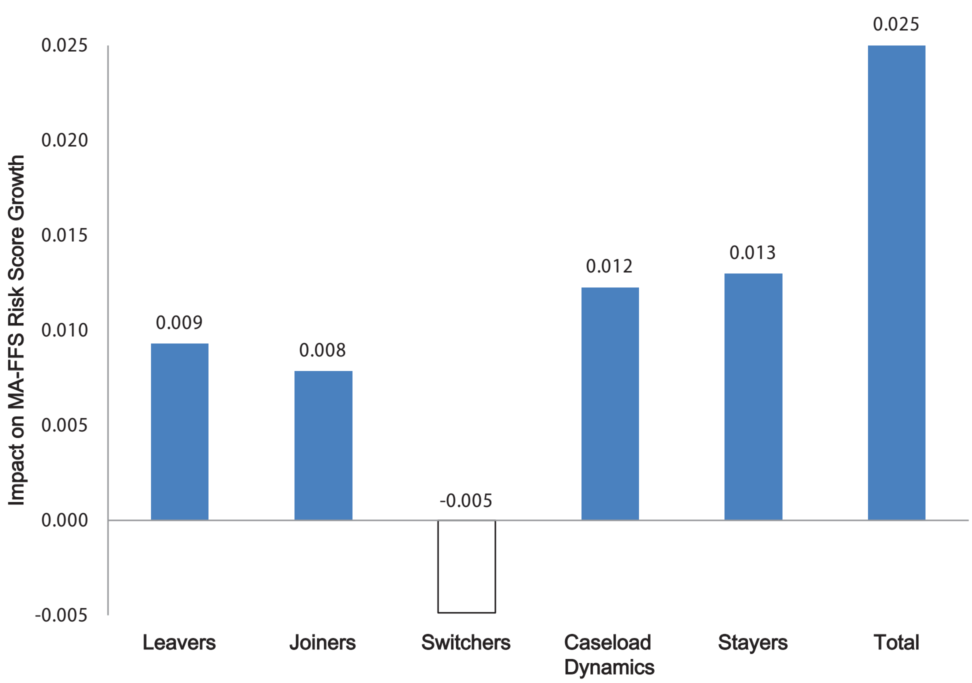

NOTE: 2004 HCC model, which here has been normalized to 1.00 for FFS in 2004. For payment purposes, this model was modified and recalibrated in 2007, 2009, 2013, and 2014. Risk scores were adjusted by -3.41%, starting in 2010. Diagnoses from FFS diagnostic radiology claims were excluded starting in 2008.
SOURCE: Authors' analysis of Medicare administrative data.

**Exhibit 2. Growth in Risk Scores Among Stayers, MA vs. FFS, 2004–2013**




NOTE: 2004 model. Stayers are beneficiaries who were enrolled in MA (or in FFS) on July 1 of two consecutive years.
SOURCE: Authors' analysis of Medicare administrative data.

**Exhibit 3. Sources of Differential Growth in MA and FFS Risk Scores, 2004–2005**



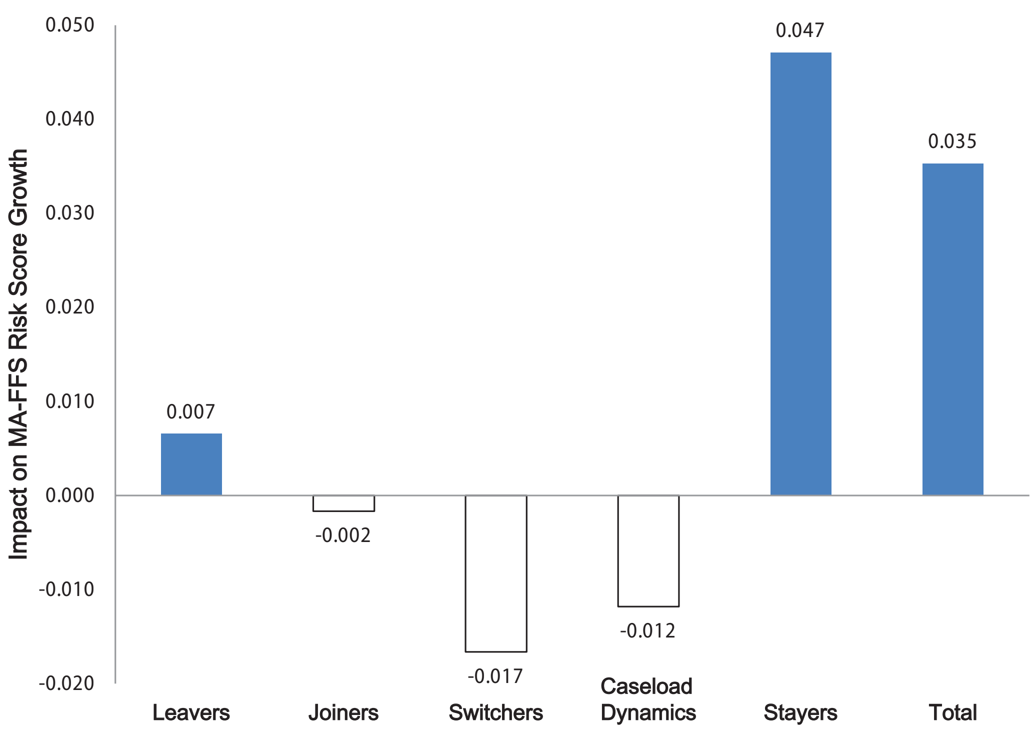

NOTE: 2004 model. Caseload dynamics is the sum of leavers (e.g., decedents), joiners (e.g., newly eligibles), and switchers. Total is the sum of caseload dynamics and stayers.
SOURCE: Authors' analysis of Medicare administrative data.

average FFS risk score increases by approximately 0.012 per year. Caseload dynamics account for the difference between the 0.088 growth for stayers and the 0.012 for all FFS beneficiaries—relatively high risk score decedents exit the caseload and are replaced by relatively low risk score 65 year olds.

From 2004 to 2005, caseload dynamics caused greater restraint in the growth of FFS risk scores than in MA (Exhibit 3). Decedents are a greater proportion of FFS beneficiaries than of MA enrollees and, similarly, 65 year olds are a greater proportion of FFS than of MA. Primarily as a result of the greater proportion of decedents in FFS, the average MA risk score in 2005 increased by 0.009 more than the average FFS score. Similarly, as a result of the greater proportion of 65 year olds in FFS, the average MA score increased by 0.008 more than the average FFS score. These effects were partially balanced by the effects of switchers—relatively low risk beneficiaries switching from FFS to MA, and relatively high risk beneficiaries switching from MA to FFS. Despite the partially counterbalancing effect of switchers, the net effect of caseload dynamics was to cause the average MA score to increase 0.012 more rapidly than the average FFS score from 2004 to 2005. The remaining 0.013 of the difference in growth rate between MA and FFS scores is accounted for by the fact that risk scores for stayers increased more rapidly in MA than in FFS.

Although caseload dynamics accounted for part of the more rapid growth in MA risk scores in 2004–2005, in 2012–2013 the effects of caseload dynamics reversed, causing the average MA risk score to grow more slowly than FFS (Exhibit 4). Decedents have a similar effect in 2012–2013 as in

**Exhibit 4. Sources of Differential Growth in MA and FFS Risk Scores, 2012–2013**




NOTE: 2004 model. Caseload dynamics is the sum of leavers (e.g., decedents), joiners (e.g., newly eligibles), and switchers. Total is the sum of caseload dynamics and stayers.
SOURCE: Authors' analysis of Medicare administrative data.

2004–2005, but the effects of joiners and switchers are different in the latter period than in the former. In contrast to the 2004–2005 results, in 2012–2013 the proportion of 65 year olds was similar in MA and FFS, and joiners do not have much effect on the difference between MA and FFS in risk score growth rates. Switchers had a small negative effect in 2004–2005, but a much larger negative effect in 2012–2013.[5] As was shown in Exhibit 2, the difference between MA and FFS in the rate of increase in risk scores for stayers widened in 2013, accounting for the larger effect of stayers in Exhibit 4 than in Exhibit 3.

On average over the 2004–2013 period, caseload dynamics had virtually no net effect on the difference between MA and FFS in the rate of growth of risk scores, and the greater increase in risk scores for MA stayers compared to FFS appears to account for most of the reason for the differential growth (Exhibit 5). Caseload dynamics led to more rapid increases in risk scores in MA in the early part of the period, but to slower increases in the latter part and, on balance, made very little contribution to the differential growth in scores.

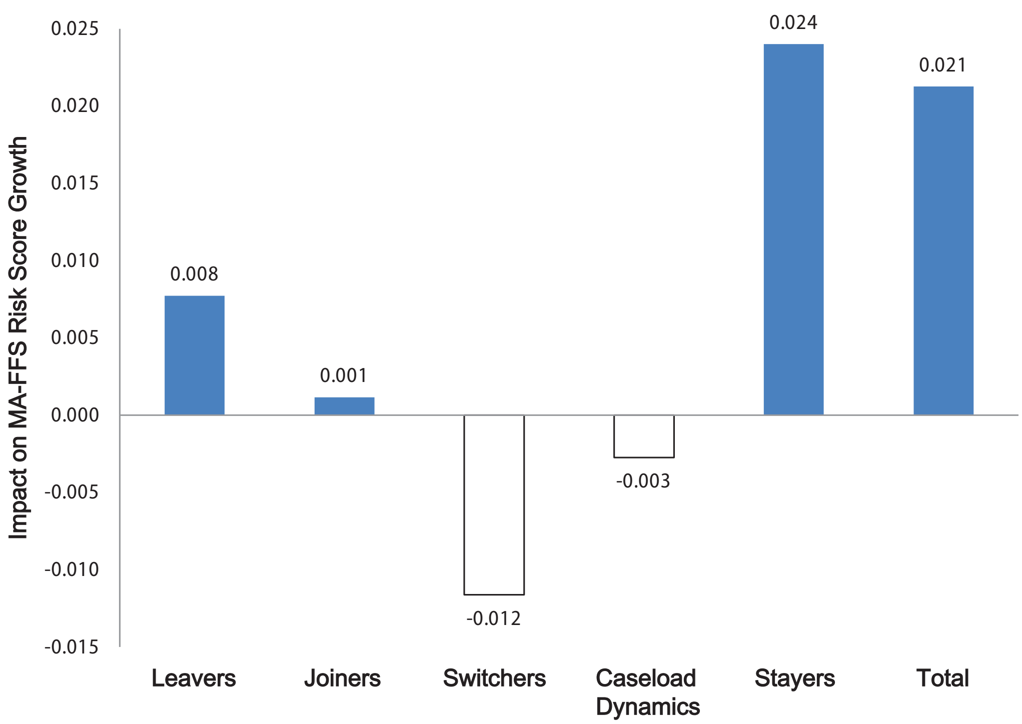
We have shown that more rapid increase in risk scores for MA stayers than for FFS stayers largely accounts for the more rapid growth in MA risk scores seen in Exhibit 1.

## Relative Morbidity of MA Enrollees: Analyses of Mortality and MCBS Data

We analyze mortality rates in MA and FFS during the 2004–2012 period to test the hypothesis that

[5] The difference between 2004-05 and 2012-13 is primarily a result of the fact that disenrollees from MA to FFS had higher risk scores in the latter cohort than in the earlier one.

**Exhibit 5. Sources of Differential Growth in MA and FFS Risk Scores, Average 2004–2013**




NOTE: 2004 model. Caseload dynamics is the sum of leavers (e.g., decedents), joiners (e.g., newly eligibles), and switchers. Total is the sum of caseload dynamics and stayers.
SOURCE: Authors' analysis of Medicare administrative data.

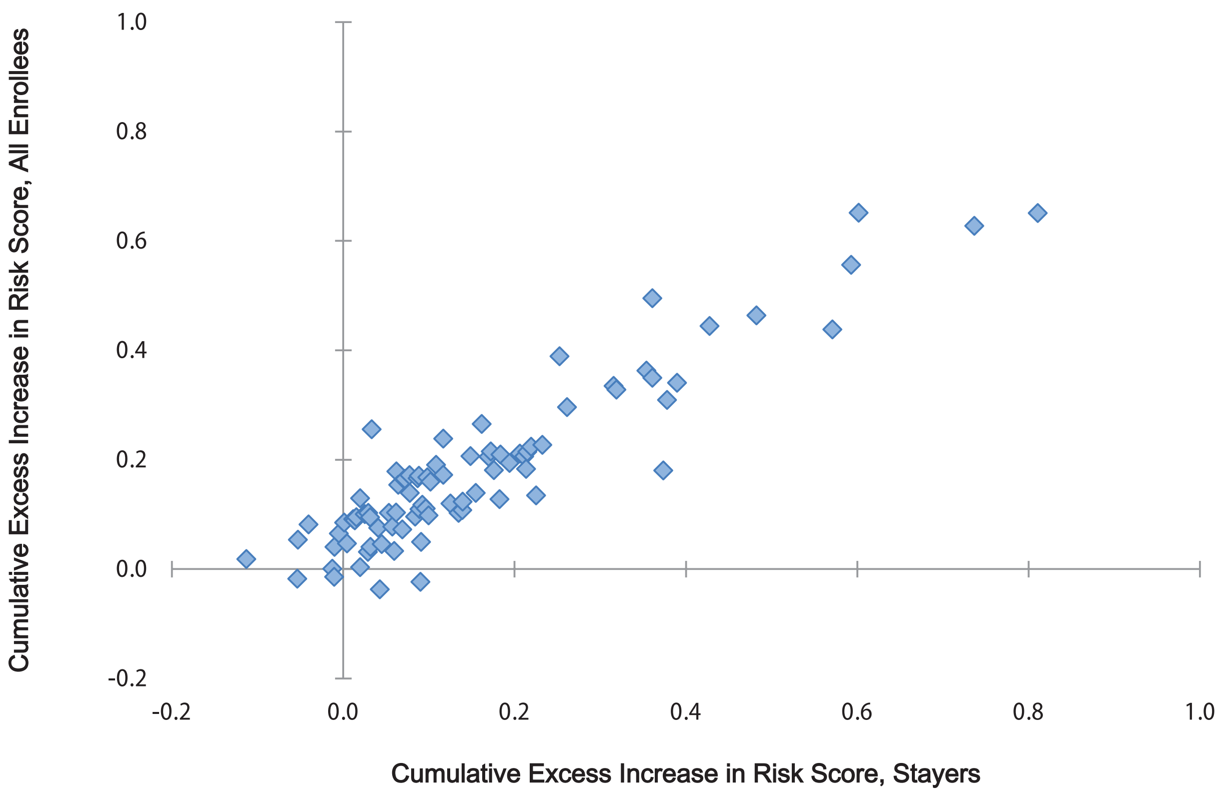
MA beneficiaries declined in health relative to FFS beneficiaries. Mortality rates declined somewhat more rapidly in MA than in FFS from 2004 to 2012, providing no support for the hypothesis that the underlying morbidity of MA beneficiaries increased relative to FFS (See Supplemental Appendix Exhibit A3 & A4). As noted earlier, mortality rates were lower in MA than in FFS.

We use MCBS data to provide an assessment of relative risk of MA and FFS beneficiaries that is independent of data reported by MA plans or FFS providers. As described in the Supplemental Appendix, we predict spending for each MA and FFS beneficiary using demographic information, self-reported health status, and the respondent's report of whether she has ever been diagnosed with hypertension, heart attack, heart failure, stroke, cancer, or diabetes. Predicted expenditures for MA enrollees average approximately 96% of predicted expenditures for FFS beneficiaries, and the ratio does not appear to have changed systematically over the 2006 to 2011 period. Results from this analysis are shown in the Supplemental Appendix Exhibit A8.

## Contract-Level Analyses

There is a striking amount of heterogeneity across MA contracts in the extent to which risk scores for stayers increased from 2004 to 2011 (Exhibit 6). The fourteen contracts accounting for 10% of MA enrollment that coded least aggressively had cumulative increases in risk scores for stayers at or below the level in FFS (i.e., cumulative excess increase at or below zero). Conversely, the four contracts in the top decile, including one with over 200,000 members, had a cumulative excess increase

**Exhibit 6. Cumulative Excess Increase in Risk Scores, All Enrollees vs. Stayers Only, by Contract, 2004–2011**




NOTE: 2004 model. Continuously participating contracts (N=86).
SOURCE: Authors' analysis of Medicare administrative data.

of 0.59 or more (at least 0.08 per year). Contracts in which risk scores for stayers increased more rapidly than in FFS also had overall risk scores that increased more rapidly than FFS (r=.92).

Increases in diagnostic coding are especially large in a relatively small number of largely discretionary HCCs (Exhibit 7).[6] For example, in 2012, 1.2% of FFS beneficiaries were coded with drug or alcohol dependence, compared to 8.1% of beneficiaries in the 10% of MA contracts with the highest level of coding intensity (as measured by the cumulative excess increase in risk scores for stayers shown in Exhibit 6). Similarly, polyneuropathy is recorded three times as often in the top decile of MA plans as in FFS. In addition, MA plans, particularly those in the top decile of coding intensity, both increased the rate at which diabetes was reported, and shifted the distribution towards the highest paying diabetes HCC (HCC 15). Among the top decile of plans, the prevalence of this HCC increased from 2.5% in 2004 to 20.1% in 2012. In contrast, for diagnoses such as heart attack and hip fracture, where there is little discretion in diagnosis, prevalence in the top decile of plans is similar to, or below, FFS levels (see Supplemental Appendix Exhibit A5).

In 2014 Medicare started using a revised model, which we term the "2014 model." In part as a result of the analysis shown in Exhibit 7, the polyneuropathy HCC was restructured to remove diabetic neuropathy, and the renal failure HCC was restructured to exclude chronic kidney disease stages 1–3 and was split into acute and chronic kidney conditions, among other changes. These changes build on the changes made in 2013 that constrained the payment weights for each of the four most severe diabetes HCCs to be equal, removing the ability of MA plans to generate extra revenue by increasing the reported severity of diabetes within these four HCCs.

As a result of the changes to the model in 2013 and 2014, the impact of coding effort on risk scores is smaller using the new 2014 model than with the older 2004 model. As shown in Exhibit 8, if the 2014 model had been in place in 2013, the average MA risk score in 2013 would have been slightly higher than the average FFS score. Even using the new model, the average MA score increased substantially relative to the average FFS score from 2004 to 2013, but the rate of increase was somewhat slower than when measured with the 2004 model—an increase of 2.2% per year in relative MA score using the 2004 model (Exhibit 1)—compared to approximately 1.6% per year using the 2014 model. This result suggests that some coding intensity efforts—such as increasing the reported severity of diabetes or coding chronic kidney disease more aggressively—have been neutralized by the 2014 model, but others have not.

## Discussion

In the 2004–2013 period, the mean MA risk score increased 2.2 percentage points per year more quickly than the mean FFS score using the 2004 model and 1.6 percentage points more quickly using the 2014 model. Comprehensive risk adjustment was first implemented in Medicare Advantage in 2004 and was not fully phased in until 2007 when 100% of Part C payments were risk adjusted. It is possible that during this phase-in period when plans were first reporting diagnosis codes from ambulatory settings, the level of coding by MA plans was not as comprehensive

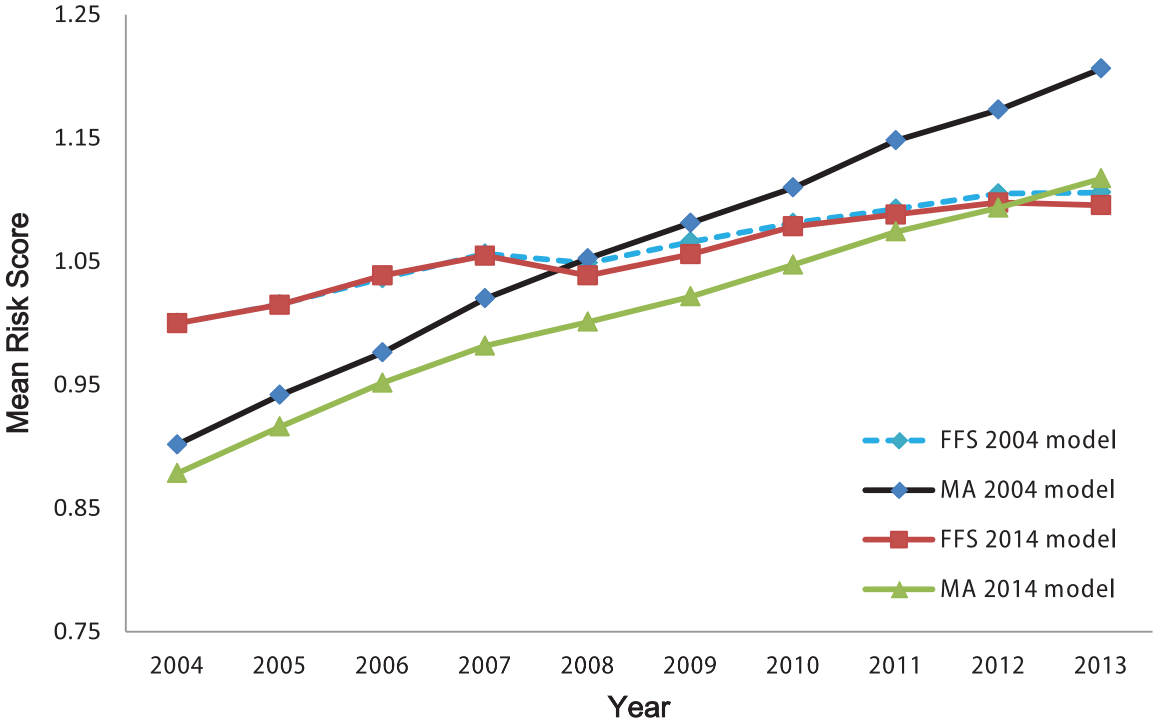
[6] An HCC is considered "discretionary" if it is based on diagnoses that providers may or may not record for patients having the diagnosis. For example, there is relatively little discretion in whether a diagnosis of 'hip fracture' should be recorded for a patient presenting with an apparent hip fracture. In contrast, there is substantial discretion about whether a patient being assessed in an annual wellness visit should be recorded as exhibiting 'alcohol or drug abuse'. Discretion might also result from either a choice of ICD-9 codes (e.g., pneumonia vs. sepsis in an inpatient setting) or underlying ambiguity of the condition (e.g., depression vs. anxiety).

**Exhibit 7. Top Ten HCCs of Relative Prevalence in the Top Decile of MA Contracts, 2004 vs. 2012.**

| | | | Prevalence | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 2004 | | | | 2012 | | | |
| Rank | HCC | Description | FFS Mean | MA Mean | MA Top Decile | Ratio, Top Decile:FFS | FFS Mean | MA Mean | MA Top Decile | Ratio, Top Decile:FFS |
| 1 | 52 | Drug/Alcohol Dependence | 0.6% | 0.4% | 0.5% | 0.75 | 1.2% | 1.7% | 8.1% | 6.64 |
| 2 | 15 | Diabetes with Renal or Peripheral Circ. Manifest. | 2.5% | 2.5% | 2.5% | 0.98 | 4.0% | 7.1% | 20.1% | 5.07 |
| 3 | 131 | Renal Failure | 3.5% | 2.7% | 2.7% | 0.79 | 9.9% | 14.6% | 38.2% | 3.85 |
| 4 | 71 | Polyneuropathy | 3.8% | 3.1% | 3.3% | 0.85 | 6.5% | 10.1% | 20.1% | 3.09 |
| 5 | 83 | Angina Pectoris/Old Myocardial Infarction | 5.3% | 4.6% | 5.0% | 0.96 | 4.9% | 7.9% | 15.0% | 3.03 |
| 6 | 55 | Major Depressive, Bipolar, & Paranoid Disorders | 3.9% | 2.4% | 3.0% | 0.76 | 5.7% | 7.5% | 14.1% | 2.50 |
| 7 | 105 | Vascular Disease | 11.2% | 7.8% | 8.1% | 0.73 | 13.6% | 15.4% | 33.3% | 2.44 |
| 8 | 108 | Chronic Obstructive Pulmonary Disease | 13.4% | 10.7% | 12.5% | 0.93 | 13.3% | 14.5% | 25.3% | 1.90 |
| 9 | 100 | Hemiplegia/Hemiparesis | 1.1% | 0.8% | 0.6% | 0.57 | 1.2% | 1.3% | 2.0% | 1.77 |
| 10 | 157 | Vertebral Fractures without Spinal Cord Injury | 1.1% | 0.8% | 0.7% | 0.62 | 1.1% | 1.0% | 1.7% | 1.58 |

NOTE: 2004 model. Low-prevalence HCCs (<1% in 2012) were excluded.
Top decile defined using 2004–2011 data.
SOURCE: Authors' analysis of Medicare administrative data.

**Exhibit 8. Mean Risk Scores, MA vs. FFS, 2004–2013 (2004 & 2014 Models)**




NOTE: CMS will use the "2014 model" for payment purposes for the first time in 2014, when it will receive a weight of 75%. This exhibit simulates past payment assuming full implementation. Here, both models have been normalized to 1.00 of the FFS mean in 2004.
SOURCE: Authors' analysis of Medicare administrative data.

as FFS. If that were the case, the relatively faster growth in risk scores during this period could be attributable to plans "catching up" to FFS, rather than coding more intensely than FFS. The true difference in average MA and FFS risk, which is beyond the scope of this paper, is a worthy topic for future research.

It appears that most of the reason that MA risk scores increased more quickly than FFS scores is due to increases in relative coding intensity—measured as increases in risk scores for stayers—with little of it accounted for by changes in enrollment mix. There is little sign of coding intensity slowing; in fact, Exhibit 2 shows that it may be increasing (using the 2004 model).

Our conclusion is supported by several findings. Our analysis of joiners, leavers, switchers, and stayers shows clearly that changes in the composition of the MA caseload do not account for the increase in relative MA risk scores. Risk scores for stayers increased more quickly for MA enrollees than for FFS beneficiaries—by an average of 0.088 for FFS stayers, compared to 0.120 for MA stayers (Exhibit 2), and it is this difference that accounts for the relative increase in MA risk scores (Exhibit 5).

In principle, risk scores for stayers could have increased more quickly in MA because of differences in characteristics at baseline; for example, risk scores typically increase more quickly for older beneficiaries than for younger beneficiaries. If MA enrollees were older than FFS beneficiaries, then MA scores (unadjusted for age) would have increased more quickly than FFS, even if there were no change in coding intensity. However, the age distribution of MA enrollees is relatively similar to the age distribution of FFS beneficiaries, and slight differences between the sectors in the age distribution do not account for differences in the rate of growth of risk scores. As a second example, risk scores may increase more quickly for beneficiaries with lower initial risk scores than for those with higher initial scores, because of regression toward the mean. However, baseline risk scores are higher in MA than in FFS in the later part of the analysis period, precluding aregression-toward-the-mean explanation for MA's more rapid growth in scores.

In addition, if MA enrollees were actually sicker, relative to FFS beneficiaries, in 2013 than in 2004, then we would have expected to see some reflection of this in our analysis of mortality rates and of self-reported health status information. In contrast, mortality rates in MA declined more quickly than mortality rates in FFS, and there was no evidence of a trend in self-reported illness, suggesting that MA enrollees did not actually become relatively sicker over the time period.

In an observational study such as this one, it is never possible to conclusively rule out all alternative hypotheses. MA risk scores increased more rapidly than FFS scores. We have shown that this did not result from a change in the composition of MA enrollees. And we have shown that using two independent measures of health status—self-reported health from the MCBS and information on mortality—that there is no indication that MA enrollees actually were getting any sicker than FFS beneficiaries. While it is theoretically possible that some explanation other than greater coding intensity accounts for the relative increase in MA risk scores, we are unable to suggest any plausible alternative hypotheses.

We have also shown that there is substantial heterogeneity across plans in coding intensity, with some plans coding at approximately the same level as FFS, while several others created a cumulative increase in coding intensity from 2004 to 2011 of 60% or more.

Coding more carefully may have real health benefits. Better identification of problems and better documentation of problems that have been identified could improve the quality of treatment provided and may even lower costs—or they may lead to unnecessary treatment and higher costs. In either case, however, the purpose and design of the risk-adjusted payment system is not to improve the quality of coding. It is to ensure that plans are paid according to the health of the patients they enroll. It is unlikely that the increased payments achieved by plans through increased coding intensity are related to substantial health benefits that better coding might produce.

CMS and the Congress have responded to the increase in risk scores over time in several ways. First, starting in 2010, CMS lowered payment by 3.41% by applying an across-the-board coding adjustment. The coding intensity adjustment will increase to 4.91% in 2014 and to at least 5.91% in 2018. Second, starting in 2013, CMS set the four most severe diabetes HCCs (HCC15-HCC18) to have the same payment coefficient (Department of Health and Human Services, 2012). As a result, recording diagnoses that move enrollees from HCC18 (diabetes with ophthalmologic or unspecified manifestation) into HCC15 (diabetes with renal or peripheral circulatory manifestation) will no longer increase revenue for MA plans. Third, CMS made further changes to the model in 2014, removing some of the HCCs that were the subject of MA efforts at increasing coding intensity.[7]

[7] CMS is also recovering overpayments identified by risk adjustment data validation (RADV) audits of selected MA contracts.

Relative MA risk scores have been increasing at least 1% per year (Exhibit 8) and are likely to continue to do so, even though MCBS-based risk scores have been roughly constant.[8] The legislated increase in coding intensity adjustment is 0.25% per year from 2014 through 2018. The President's Budget for 2015 proposes increasing the minimum adjustment to 0.67% per year through 2020, plateauing at 8.51% in 2020 and thereafter, much closer to the expected increase in relative risk scores. The across-the-board adjustments do not address the substantial heterogeneity in plan-level results shown in Exhibit 6.

It is challenging to accurately measure the effects of coding intensity on MA risk scores and even more challenging to devise optimal policy responses. Uncertainty about the extent to which coding intensity would persist beyond the early years of using diagnostic information, uncertainty about whether some of the changes observed in MA risk scores might have been due to changes in underlying morbidity of MA enrollees, and heterogeneity across plans in coding intensity each create challenges in evaluating appropriate policy responses. In addition, it is unclear to what extent MA plans coded less completely than FFS in 2004, and how much of the increase in coding intensity reflects MA simply catching up to FFS levels, and how much represents going above and beyond. The Department of Health and Human Services is continuing its analysis of the relative risk of MA and FFS enrollees to improve the ability of the Medicare program to accurately pay for MA beneficiaries.

## Disclaimer

The authors report no conflicts of interest related to this analysis. The content is solely the responsibility of the authors and does not necessarily reflect the views of the Department of Health and Human Services.

[8] Some would expect that MA plans will react to the 2013 and 2014 model changes by finding other HCCs on which to focus their efforts, and the success of coding intensity efforts may well increase in the future.

## Correspondence

W. Pete Welch, Ph.D., Department of Health and Human Services—Assistant Secretary for Planning and Evaluation (ASPE), 200 Independence Ave., SW, Washington, DC 20201, Pete.Welch@hhs.gov, Tel. 202-260-6213

## Acknowledgment

Data were available through a data use agreement (#212990) between the Office of the Assistant Secretary for Planning and Evaluation (ASPE) and the Centers for Medicare & Medicaid Services (CMS). The authors appreciate the assistance of David Freund of Fu Associates and Andrew Bindman and the programming work of Acumen, LLC.

# References

Brown, R. S., Bergeron, J. W., Clement, D. G., Hill, J. W., & Retchin, S. (1993). Does Managed Care Work for Medicare? An Evaluation of the Medicare Risk Program for HMOs. Princeton, NJ: Mathematica Policy Research.

Call, K. T., Dowd, B., Feldman, R., & Maciejewski, M. (1999). Selection Experiences in Medicare HMOs: Pre-Enrollment Expenditures. *Health Care Financing Review*, *20*(4), 197–209. PubMed

Department of Health and Human Services (2007). Advance Notice of Methodological Changes for Calendar Year 2008 for Medicare Advantage (MA) Capitation Rates.

Department of Health and Human Services (2009). Advance Notice of Methodological Changes for Calendar Year 2010 for Medicare Advantage (MA) Capitation Rates.

Department of Health and Human Services (2012). Advance Notice of Methodological Changes for

Calendar Year 2013 for Medicare Advantage (MA) Capitation Rates.

Department of Health and Human Services (2013). Announcement of Calendar Year (CY) 2014 Medicare Advantage Capitation Rates and Medicare Advantage and Part D Payment Policies and Final Call Letter.

Government Accountability Office. (2012). Medicare Advantage: CMS Should Improve the Accuracy of Risk Score Adjustments for Diagnostic Coding Practices. *Report to Congressional Requesters* (GAO-12-51). Retrieved from the U.S. GAO Web site: http://www.gao.gov/products/GAO-12-51

Gorman Health Group (2013). Retrieved from https://www.gormanhealthgroup.com/#all

Hellinger, F. J., & Wong, H. (2000). Selection Bias in HMOs: A Review of the Evidence. *Medical Care Research and Review*, *57*(4), 405–439. PubMed http://dx.doi.org/10.1177/107755870005700402

Leprechaun. (2013). Retrieved from http://www.lepmed.com/retrospective_chart_review.asp

Lichtenstein, R., Thomas, J. W., Adams-Watson, J., Lepkowski, J., & Simone, B. (1991). Selection Bias in TEFRA At-Risk HMOs. *Medical Care*, *29*(4), 318–331. PubMed http://dx.doi.org/10.1097/00005650-199104000-00002

McWilliams, J. M., Hsu, J., & Newhouse, J. (2012). New Risk-Adjustment System Was Associated With Reduced Favorable Selection in Medicare Advantage. *Health Affairs*, *31*(12), 2630–2640. PubMed http://dx.doi.org/10.1377/hlthaff.2011.1344

Medicare Payment Advisory Commission (1998). Report to the Congress: Medicare Payment Policy. Washington, DC.

Medicare Payment Advisory Commission. (2012, June). Issues for Risk Adjustment in Medicare Advantage. In *Report to the Congress: Medicare and the Health Care Delivery System* (chapter 4). Washington, DC

Mello, M. M., Stearns, S. C., Norton, E. C., & Ricketts, T. C., III. (2003). Understanding Biased Selection in Medicare HMOs. *Health Services Research*, *38*(3), 961–992. PubMed http://dx.doi.org/10.1111/1475-6773.00156

Morrisey, M. A., Kilgore, M. L., Becker, D. J., Smith, W., & Delzell, E. (2012, October). Favorable Selection, Risk Adjustment, and the Medicare Advantage Program. *Health Services Research*, *48*(3), 1036–1039.

Newhouse, J. P., Price, M., Huang, J., McWilliams, J. M., & Hsu, J. (2012). Steps To Reduce Favorable Risk Selection In Medicare Advantage Largely Succeeded, Boding Well For Health Insurance Exchanges. *Health Affairs*, *31*(12), 2618–2628. PubMed http://dx.doi.org/10.1377/hlthaff.2012.0345

PPACA & HCERA (Patient Protection and Affordable Care Act & Health Care and Education Reconciliation Act). 2010.

Physician Payment Review Commission (1996). Risk Selection and Risk Adjustment in Medicare. In the *Annual Report to Congress* (chapter 15).

Pope, G. C., Kautter, J., Ellis, R. P., Ash, A. S., Ayanian, J. Z., Iezzoni, L. I., . . . Robst, J. (2004). Risk Adjustment of Medicare Capitation Payments using the CMS-HCC Model. *Health Care Financing Review*, *25*(4), 119–141. PubMed

Riley, G. F. (2012). Impact of Continued Biased Disenrollment from the Medicare Advantage Program to Fee-for-Service. *Medicare & Medicaid*

*Research Review*, *2*(4), E1–E17. PubMed http://dx.doi.org/10.5600/mmrr.002.04.a08

Riley, G., Chiang, Y. P., Ingber, M. J., & Tudor, C. G. (1996). Health Status of Medicare Enrollees in HMOs and Fee-For-Service in 1994. *Health Care Financing Review*, *17*(4), 65–76. PubMed

Welch, H. G., Sharp, S. M., Gottlieb, D. J., Skinner, J. S., & Wennberg, J. E. (2011). Geographic Variation in Diagnosis Frequency and Risk of Death among Medicare Beneficiaries. *Journal of the American Medical Association*, *305*(11), 1113–1118. PubMed http://dx.doi.org/10.1001/jama.2011.307