

## How Does Risk Selection Respond to Risk Adjustment? New Evidence from the Medicare Advantage Program<sup>†</sup>

By JASON BROWN, MARK DUGGAN, ILYANA KUZIEMKO,  
AND WILLIAM WOOLSTON\*

*To combat adverse selection, governments increasingly base payments to health plans and providers on enrollees' scores from risk-adjustment formulae. In 2004, Medicare began to risk-adjust capitation payments to private Medicare Advantage (MA) plans to reduce selection-driven overpayments. But because the variance of medical costs increases with the predicted mean, incentivizing enrollment of individuals with higher scores can increase the scope for enrolling "overpriced" individuals with costs significantly below the formula's prediction. Indeed, after risk adjustment, MA plans enrolled individuals with higher scores but lower costs conditional on their score. We find no evidence that overpayments were on net reduced. (JEL G22, H51, I13, I18)*

Recent health care reforms have attempted to move away from the fee-for-service (FFS) payment model—which economists have long argued incentivizes over-provision of services—by paying providers or insurers fixed capitation payments rather than reimbursing them for each service. The success of such reforms hinges on correctly aligning capitation payments with a patient's expected cost. Otherwise, plans and providers will have incentives to cream skim overpriced cases instead of competing on quality or cost.

To more accurately equate payments with expected costs, governments and other insurance sponsors have increasingly turned to "risk adjustment"—setting payments to insurers or providers to take account of an individual's past and current health conditions. For example, the Affordable Care Act of 2010 (ACA) relies heavily on

\* Brown: US Treasury, Office of Economic Policy, 1500 Pennsylvania Avenue NW, Washington, DC 20005 (e-mail: Jason.Brown@do.treas.gov); Duggan: Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94304 (e-mail: mgduggan@stanford.edu); Kuziemko: Department of Economics, Princeton University, 322 Wallace Hall, Princeton, NJ 08544 (e-mail: kuziemko@princeton.edu); Woolston: J.P. Morgan Chase, 270 Park Avenue, New York, NY 10017 (e-mail: william.woolston@gmail.com). We thank Richard Boylan, Doug Bernheim, Sean Creighton, David Cutler, Liran Einav, Randy Ellis, Zeke Emanuel, Gopi Shah Goda, Jonathan Gruber, Caroline Hoxby, Seema Jayachandran, Larry Katz, Robert Kocher, Jonathan Kolstad, Amanda Kowalski, Alan Krueger, Joe Newhouse, Elena Nikolova, Christina Romer, Shanna Rose, Karl Scholz, Jonathan Skinner, Luke Stein, and seminar participants at Brown, Columbia, Cornell, Harvard, Houston, Northwestern, MIT, Princeton, RAND, Rice, Stanford, USC, US Treasury Department, Wharton, Wisconsin, Yale, and the NBER Health Care meetings for helpful comments and feedback. We thank Nadia Tareen and Boris Vabson for their outstanding research assistance. Views expressed here are solely those of the authors and not of the institutions with which they are affiliated, and all errors are our own.

<sup>†</sup> Go to <http://dx.doi.org/10.1257/aer.104.10.3335> to visit the article page for additional materials and author disclosure statement(s).

risk adjustment.<sup>1</sup> However, empirical research on these attempts to risk-adjust has been limited.<sup>2</sup>

In this paper, we provide an assessment of the largest risk-adjustment effort to date in the US health care sector—Medicare’s risk adjustment of capitation payments to private Medicare Advantage (MA) plans, which the ACA suggests as the model for risk adjustment in the state-run insurance exchanges—on selection into MA plans and on the government’s total cost of financing Medicare benefits. Since the 1980s, Medicare enrollees have been able to enroll in either the traditional fee-for-service (FFS) program or in an MA plan, which can provide additional services but must cover the basic benefits guaranteed by traditional Medicare. For an individual in an MA plan, the government pays the plan a capitation payment meant to cover the cost of providing her Medicare benefits. Today, more than one-fourth of Medicare’s 52 million enrollees receive their care through a private MA plan.

Before 2004, an MA enrollee’s capitation payment was, essentially, based on the average cost of FFS enrollees with the same demographic characteristics in her county and was not adjusted for health conditions. Despite regulations requiring MA plans to offer the same plan at the same price to all Medicare beneficiaries in its geographical area of operation, researchers found that less costly individuals were much more likely to enroll in an MA plan.<sup>3</sup> Reacting to this evidence of “differential payments” to MA plans—payments in excess of the expected cost of covering a beneficiary in traditional FFS—in 2004 Medicare began to base capitation payments on an individual’s “risk score,” generated by a risk-adjustment formula accounting for more than 70 disease conditions.

We develop a simple model to show that plans’ endogenous response to risk adjustment can undo the intended goal of reducing overpayments and test it using data from the Medicare Current Beneficiary Survey (MCBS). Before risk adjustment, MA plans had an incentive to enroll individuals who were low cost on all dimensions. After risk adjustment, plans no longer need to avoid beneficiaries with conditions *included in the formula*. In a difference-in-differences model, we show that, relative to individuals who remain in FFS, risk scores of those joining MA increase after risk adjustment, consistent with our model’s predictions.

However, our model emphasizes how selection can take place on different margins. While risk adjustment indeed decreases plans’ scope for advantageous selection along the dimensions *included* in the formula, it increases the incentive to find individuals who are positively selected along dimensions *excluded* from the formula and are thus “cheap for their risk score.” Indeed, as the model predicts, we find that

<sup>1</sup> Approximately 25 million people are projected to join the insurance exchanges established by the ACA, in which private insurers will receive capitation payments adjusted for enrollees’ health status (Congressional Budget Office 2014).

<sup>2</sup> There is a large, mostly theoretical or statistical, literature on risk adjustment, and Van de Ven and Ellis (2000) and Ellis (2008) serve as excellent reviews. Recently, work has focused on “optimal” risk adjustment, following Glazer and McGuire (2000) who argue that mere predictive models (such as the one used by Medicare, on which we focus the empirical work) are fundamentally misguided because formula coefficients need to be chosen for their incentive, not predictive, properties. However, as noted by Ellis (2008), predictive models are by far the most common risk adjustment models in use today, and thus determining their effect on selection and costs is a central policy question. On the empirical side, Bundorf, Levin, and Mahoney (2012) provides estimates on the welfare gains to risk adjustment of health insurance *premiums*.

<sup>3</sup> See, e.g., Langwell and Hadley (1989); Physician Payment Review Commission (1997); Mello et al. (2003); and Batata (2004).

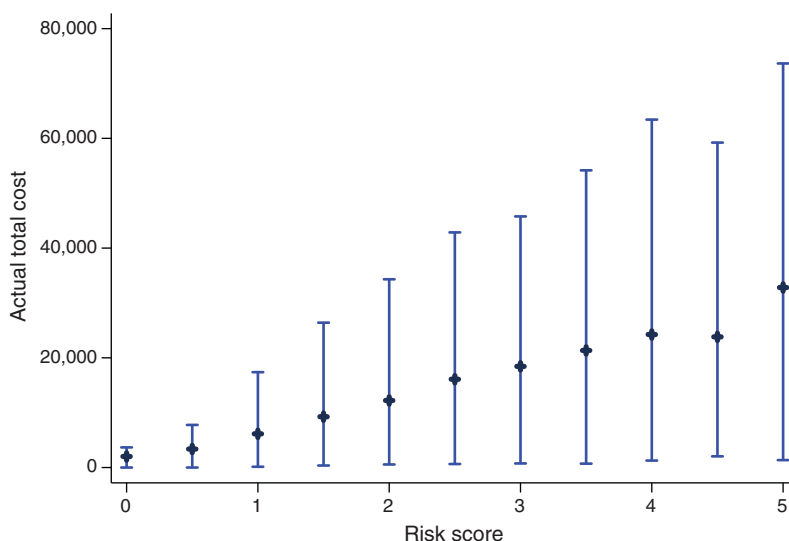


FIGURE 1. MEANS, 10TH, AND 90TH PERCENTILES OF TOTAL COSTS, BY RISK SCORE

*Notes:* All observations spent all 12 months of the previous year in FFS (so that current-year risk scores can be calculated) and no months of the current year in MA (so that all current cost data can be observed). Observations are taken only from the pre-period (before 2004) so that the sample is unlikely to be selected with respect to the risk score. Sample weights provided by the MCBS are used.

actual costs *conditional on the risk score* of those joining MA fall substantially after 2003, relative to those remaining in FFS.

Finally, the model makes clear that the former effect (the decrease in selection along dimensions included in the formula) can be more than offset by the latter effect (the increased selection conditional on the risk score). The key insight is that because the variance of medical costs increases with the expected mean, there are more cases of extremely high overpayments among those with high risk scores. Figure 1, which plots average medical costs along with the 10th and 90th percentile, shows how the variance of medical costs increases with a patient's risk score. Given that costs are bounded below by zero, overpayments to those with a risk score of 0.5 are bounded above by \$3,500, whereas if plans can manage to avoid the costliest 10 percent of enrollees with a risk score of two (five), their overpayments for this group would average over \$5,000 (\$9,000).<sup>4</sup>

Due to this increase in variance, the ability of firms to enroll individuals with costs substantially below the formula's prediction—whether through targeted advertising or designing benefits packages that differentially appeal to certain people—can actually increase after risk adjustment, and with it the government's total cost of financing the Medicare program. To take but one example from our data, pre-risk-adjustment, Hispanics were roughly \$1,200 cheaper on average than their (non-risk-adjusted) capitation payments; after risk adjustment, Hispanics with a history of congestive heart failure (one of the most common conditions included in the risk formula) are on average \$3,500 cheaper than their (risk-adjusted) capitation

<sup>4</sup>For these calculations, we use average FFS costs as an estimate for plan per-enrollee payments. The next section discusses modifications to this formula over time.

payments. Intuitively, before risk adjustment MA plans fished in a pond of relatively healthy enrollees with little cost variance. Risk adjustment allows them to fish in a pond of enrollees who have higher costs on average but also highly *variable* costs. Indeed, we find that after risk adjustment, overpayments are higher, an increase equal to roughly 9 percent of average Medicare per capita spending.

This counterintuitive consequence of risk adjustment has, to the best of our knowledge, not been noted by other researchers, but is related to the literature on the unintended consequences of increasing the specificity of incomplete contracts. By selecting individuals with low costs conditional on their risk scores, MA firms' behavior is analogous to the worker who focuses on the contractable task to the detriment of other tasks (as in Holmstrom and Milgrom 1991) or the instructor who "teaches to the test" at the expense of other educational goals (as in Lazear 2006). More generally, our results suggest that using additional information to determine prices can sometimes aggravate problems associated with asymmetric information, as in Einav and Finkelstein (2011).

While we find little evidence that risk adjustment accomplished the goal of reducing overpayments, we also examine whether the increased overpayments led to greater consumer or producer surplus. Our results suggest little to no improvement in several alternative measures of beneficiary satisfaction and quality of care. These results are consistent with recent research regarding the incidence of MA reimbursement generosity (Cabral, Geruso, and Mahoney 2014), with additional advertising expenditures absorbing much of the additional Medicare spending (Mehrotra, Grier, and Dudley 2006 and Duggan, Starc, and Vabson 2014). Perhaps because of these additional marketing costs, benefits to plans were also limited, with CMS actually increasing plan reimbursement to cushion the expected negative effect of risk adjustment on insurers' profits.

The remainder of the paper is organized as follows. Section I provides background information on the MA program and the risk-adjustment formula Medicare currently uses. Section II presents the intuition and results from the model. Section III describes the data. Sections IV and V present the empirical results on selection and differential payments, respectively. Section VI explores potential mechanisms by which MA plans might be able to differentially select certain enrollees. Section VII explores the welfare consequences of risk adjustment and discusses ways to improve it, and Section VIII concludes.

## **I. Medicare Advantage Capitation Payments and Risk Adjustment**

Since the 1980s, Medicare enrollees have had the choice between the traditional FFS program and private MA plans (previously known as Medicare+Choice plans). The evolution of MA enrollment as a share of total Medicare enrollment during our sample period is plotted in online Appendix Figure 1.

Plans must accept all applicants residing in their areas of operation and provide benefits that are covered under traditional Medicare. MA plans have considerable latitude in creating their hospital and physician networks. Many offer extra benefits such as vision care, dental care, and gym memberships. Plans can also charge a monthly premium, reduce enrollees' Medicare Part B premiums, or vary copayments.

The Medicare program pays MA plans a fixed capitation payment to cover these costs (excluding hospice care, which FFS covers), and plans are, essentially, the residual claimants if actual costs are above or below the capitation payment. Since 2006, Medicare Part D has provided enrollees coverage for prescription drugs, though all of our analysis will focus on Part A (hospital and inpatient) and B (physician and outpatient), as these are the services MA plans are required to provide.<sup>5</sup>

The capitation payment to an MA plan for covering an individual is based on the estimated Part A and B payments had FFS Medicare covered her directly. During the 1980s and 1990s, the Center for Medicare and Medicaid Services (CMS)—the agency that administers Medicare—used a “demographic model” to perform this estimation, so-called because it included primarily demographic variables (gender, age, and disability, Medicaid and institutional status) as opposed to disease or health conditions. The demographic model would output a “risk score” (with mean one) that when multiplied by a county-level “benchmark” would determine the capitation payment. Then as now, CMS did not require MA plans to report cost or claims data, so it used FFS data to regress total Part A and B spending on these demographic factors, finding that 1 percent of FFS expenditures were explained by the risk score (Pope et al. 2004).

In response to research showing that MA plans enrolled beneficiaries who were significantly cheaper than the demographic model predicted, CMS revised its risk-adjustment procedure.<sup>6</sup> In 2000, CMS made 10 percent of capitation payments dependent on inpatient claims data, raising the effective  $R^2$  of the formula from 1.0 to 1.5 percent. More significantly, in 2004—which for simplicity we term the “start” of risk adjustment—CMS introduced the hierarchical condition categories (HCC) model, still in use today. The HCC model, like the demographic model, uses data from the FFS population to predict FFS costs in the following year, but instead of relying only on demographic data, it also accounts for the disease conditions included on FFS providers’ claims. The model distills the roughly 15,000 ICD-9 codes that providers can list on claims into 70 disease-category indicator variables, the most common of which are described in online Appendix Table 1. By definition, these variables are the same whether a person has 1 or 100 claims for a certain condition. Initially, the HCC model was blended with the demographic model, and accounted for 30, 50, 75, and 100 percent of the total risk score in, respectively, 2004, 2005, 2006, and 2007 or later. To help plans adjust to the new system, CMS increased payments across the board to MA plans after risk adjustment (we discuss the potential effects of these payments in the next section).

CMS found that within the FFS population, the HCC risk score explained 11 percent of FFS expenditures the following year (Pope et al. 2004). Newhouse, Buntin, and Chapman (1997) and Van de Ven and Ellis (2000) survey the literature and conclude that the lower bound on the percent of expenditure variation that insurers are able to predict is between 20 and 25 percent, suggesting there is still room for risk selection even if the model performs as well on the MA population as it does on the

<sup>5</sup> MA plans that provide prescription drug coverage receive a separate capitation payment in return.

<sup>6</sup> Estimates from Langwell and Hadley (1989); Physician Payment Review Commission (1997); Mello et al. (2003); and Batata (2004) suggest that individuals switching from traditional FFS to MA had medical costs between 20 and 37 percent lower than observably similar individuals who remained in FFS.

FFS population. Similarly, reports commissioned by CMS in 2000 and 2004 (Pope et al. 2000 and Pope et al. 2004) and more recent work (Frogner et al. 2011) have found that—again, looking only at the FFS population—the formula systematically under-predicts spending for those with the most serious health conditions.

It is worth noting that, for at least three reasons, the spending prediction from the HCC model is likely to perform worse on the MA population than on those in FFS. First, out-of-sample prediction is more difficult than in-sample prediction. Second, CMS has found that MA plans exhibit greater “coding intensity” in documenting disease conditions than do FFS providers. For example, what an FFS provider codes as “diabetes” an MA plan might code “diabetes with complications,” thus increasing the enrollee’s capitation payment (CMS 2010). Third, MA plans may target beneficiaries for whom the formula over-predicts costs. Indeed, as the model in the next section demonstrates, risk adjustment incentivizes insurers to enroll individuals whom they expect to have low costs *conditional on their risk score*.

## II. Theoretical Framework

Our model of how plans will respond to risk adjustment relies on a simple, under-appreciated fact about medical costs: as its expectation rises, so does the variance around that expectation. One might paraphrase and say that healthy people are all alike, but sick people are each sick in their own way.

Before risk adjustment, when plans were roughly getting about \$8,000 per enrollee, regardless of medical history, it made little financial sense for a plan to enroll someone with a risk-score of, say, five, meaning expected costs of \$40,000. Yet, as Figure 1 shows, because of the substantial variance such an individual exhibits, post risk-adjustment the margin between the capitation payment for these individuals and actual costs can be significant if plans can engage in even modest risk selection. As noted in the introduction, merely avoiding the costliest 10 percent of enrollees within a risk group nets substantial margins (e.g., roughly \$7,500 on average for those with risk scores between three and four).

The idea of large potential margins among those with high risk scores is captured nicely in the below quote from Thomas Scully, the director of CMS from 2001 to 2003 and currently a general partner in a private equity firm focusing on health care:

*If you get paid \$10,000 per year for everybody [as in the pre-risk-adjustment regime], you are going to find healthy people and avoid the sick people. Well, now we have risk adjustment in Medicare... [Insurance plans] want to find a \$50,000 patient because ... you can't make an \$8,000 margin when Medicare is paying you \$8,000. Risk adjustment has totally flipped all of the incentives in Medicare for insurance companies.<sup>7</sup>*

The quote emphasizes that, perhaps ironically, the margin between capitation payments and medical costs (i.e., differential payments) can actually increase post-risk adjustment, now that plans can “fish” in a high-variance, high-expected-mean “pond.”

<sup>7</sup>From an April 28, 2011 conference at Columbia Business School. A full video of his remarks can be found at [http://www.youtube.com/watch?v=r\\_agpZpHhh8&feature=relmfu](http://www.youtube.com/watch?v=r_agpZpHhh8&feature=relmfu) (accessed September 5, 2014).



TABLE 1—EXAMPLE ILLUSTRATING MODEL PREDICTIONS

		Has cancer	
	No conditions	In remission	In treatment
<i>Model fundamentals</i>			
True costs	5	6	13
Screening	1	2	2
Risk score	5	9.5	9.5
Residual (cost – r. score)	0	–3.5	3.5
<i>Not risk adjusted</i>			
Capitation payments	8	8	8
Differential payments	3	2	–5
Profits	2	0	–7
<i>Risk adjusted</i>			
Capitation payments	5	9.5	9.5
Differential payments	0	3.5	–3.5
Profits	–1	1.5	–5.5

Note: Boxes indicate the type of enrollee that will join MA under each regime.

### A. Illustrating the Theory with a Simple Example

We begin with a simple three-type example that can show all the key insights of the model, and then describe how these results are generalized in the online mathematical Appendix.

*Basic Setup.*—There are three types of individuals, one of whom is “healthy” and two of whom are “sick.” Each type represents a third of the population. To fix ideas, type A is “healthy” and has no documented health conditions. He has expected costs of 5 were he to be covered directly by FFS, and there is no cost variation within members of type A: recall, healthy people are all alike. “Sick” types have cancer, and are not all alike. Type B has cancer that is in remission and has costs of 6; type C is receiving chemotherapy and has costs of 13. Table 1 displays this information. For simplicity, we assume that medical costs of treating each type is the same in FFS as for an MA plan.<sup>8</sup>

Before risk adjustment, capitation payments are set equal to average FFS costs across all types, or  $8 \left( \frac{5+6+13}{3} \right)$ . After risk adjustment, the government pays plans the average cost in each *risk category*—that is, no conditions (type A) and cancer (types B and C). For simplicity, we assume that the risk score is equal to average cost, so the risk score for those with no conditions (type A) equals 5 and the risk score for those with cancer (types B and C) equals  $\left( \frac{6+13}{2} \right) = 9.5$ . Note that risk adjustment is “payment-neutral” in the sense that, if the entire Medicare population joined MA, total capitation payments would be the same before and after risk adjustment,  $3 \times 8 = 24$  and  $5 + 9.5 + 9.5 = 24$ , respectively.

While MA plans must accept any individual who wishes to join, we assume that a plan can—at some cost—influence the characteristics of its enrollees. These

<sup>8</sup>In practice, MA plans may affect the utilization of health care or negotiate different prices with providers.

screening costs could include targeted advertising or additional benefits that appeal to certain groups. We assume that it is costlier to screen *within* a risk category than *across* categories. Applying these assumptions to our example, we set screening costs for type A individuals at just 1 while costs for type B (or C) individuals are 2. Put another way, it costs less to attract *generally* healthy individuals than to attract *relatively* healthy cancer patients. However, *the cost of screening within a risk category falls with the category's mean cost*. In our example, the cost of screening among type A individuals is infinite, as “healthy people are all alike.” The cost of screening among those in the cancer risk category is positive but finite. If, instead, firms decline to influence their enrollment and merely open their door to all comers, screening costs are assumed to be zero.

*Enrollment Before Risk Adjustment.*—Profits are defined as capitation payment less medical and screening costs (if any). If plans do not screen, they make zero expected profits, as capitation payments and medical costs are equal in expectation. If they screen, as Table 1 shows, profits pre-risk-adjustment are 2, 0, and  $-7$  when plans selectively enroll types A, B, and C, respectively. We assume that insurers only enroll profitable individuals, so plans will choose to selectively enroll type A individuals and earn profits of 2.

*Enrollment After Risk Adjustment.*—Again, if plans do not screen, they make zero expected profits because both pre- and post-risk adjustment, medical costs and capitation payments are equal in expectation. If they screen, profits are  $-1$ ,  $1.5$ , and  $-5.5$  for selectively enrolling types A, B, and C, respectively. Plans thus enroll type B and earn profits of  $1.5$ .

*Results.*—The first outcome of note is that the *risk scores of those enrolled in MA increase after risk adjustment*, specifically from 5 (A's risk score) in the pre-period to  $9.5$  (B's risk score) in the post-period. Intuitively, post risk adjustment there is no longer a penalty for enrolling individuals with high risk scores, so plans no longer expend the screening costs to avoid such individuals. We term selectively enrolling those with low risk-scores “extensive-margin screening” and our model thus predicts that it falls after risk adjustment.

The second outcome of note is that *for those enrolled in MA, medical costs conditional on the risk score fall after risk adjustment*. Specifically, in the pre-period, medical costs less the risk score were  $5 - 5 = 0$ , falling to  $6 - 9.5 = -3.5$  in the post-period. The key to this result is that within-risk-score screening costs fall with the risk score itself. To paraphrase the quote, it is impossible to find someone with a \$10,000 margin when they have a low risk score and thus, say, an \$8,000 capitation payment. But, for patients with high risk scores, such a margin is possible because of the high variance. Again, in our example, the cost of screening within type A (*risk score* = 5) is infinite, as no variation exists.<sup>9</sup> We term selectively enrolling those with costs below their risk score “intensive-margin selection.”

<sup>9</sup>We make this assumption so as not to need four types to illustrate the point, but as we show in the online Appendix, the result of increasing intensive-margin selection does not depend on the particular screening costs we choose here.



A third outcome is that *risk adjustment would have reduced differential payments had the population joining MA remained fixed*. In our example, only type A joins MA before risk adjustment, and risk adjustment reduces differential payments for this population to zero.

A fourth outcome from our example is that the government's differential payments (capitation payments less medical costs) actually rise under risk adjustment. In the pre-period, differential payments were  $8 - 5 = 3$ , rising to  $9.5 - 6 = 3.5$  in the post-period.

A fifth outcome is that profits fall. Because risk adjustment changes the screening costs that insurers pay, differential payments and plan profits need not move together. In our original example, profits actually fall from 2 to 1.5 given the increased screening costs. As we do not have data on MA-specific insurer profits, we cannot directly test this result, but we return to it when we discuss welfare in Section VII.<sup>10</sup>

In the online Appendix, we go from three types to a continuum of types and allow the predictive power of risk adjustment to vary continuously as well. As we show, all five results from our three-type discrete set-up hold, with an important exception: the effect of risk adjustment on overpayments is ambiguous. The reason for the ambiguity reinforces the main theme of the paper: the success of risk adjustment depends crucially on how much medical cost variance increases with the risk score. Suppose in our example that instead of ranging from 6 to 13, the costs of those with cancer range only from 7 (type B) to 12 (type C). In the post-period, it is still the case that plans only enroll type B. Our two selection results hold: for those in MA, risk scores rise from 5 to 9.5 ("extensive-margin" selection falls) and costs less the risk score falls, in this case from  $5 - 5 = 0$  to  $7 - 9.5 = -2.5$  ("intensive-margin" selection increases). However, risk adjustment in this case has accomplished its goal of reducing the government's differential payments, from 3 in the pre-period to  $9.5 - 7 = 2.5$  in the post-period.

### B. Discussion of Assumptions

First, a central though seemingly innocuous assumption of the model is *payment neutrality*, that if MA plans were to enroll all Medicare enrollees (or a random sample thereof), total payments would be the same before and after risk adjustment. If, instead, risk adjustment is accompanied by an increase in what we term "statutory" overpayments—that is, overpayment related to the government's decision to systematically overpay MA plans on average, even absent risk-selection—then our predictions need not hold. Suppose that along with risk adjustment, Medicare decided to increase all capitation payments by 20 percent, as we illustrate in online Appendix Table 2. Results are unchanged in the pre-period, but now plans can either engage in screening, in which case it is most profitable to differentially enroll type B for a profit of  $9.5 \times 1.2 - 6 - 2 = 3.4$ . Or, they can open their doors to all comers and gain  $1.2 \times (5 + 9.5 + 9.5) - 5 - 6 - 13 = 4.8$ , as they incur no screening

<sup>10</sup>Finally, we note that our model yields ambiguous predictions on how the average medical costs of MA enrollees should change. On the one hand, after risk adjustment MA enrollees will have higher risk scores. On the other hand, their costs conditional on their score will fall. In the online Appendix, we show that the second effect can dominate and that risk adjustment can cause average costs of MA enrollees to fall. We hasten to add that this result is not general—indeed, in our example medical costs among those in MA increase from 5 to 6—but is possible.

costs. As such, if risk adjustment is accompanied with large increases in statutory overpayments, plans will at some point lose any incentive to find those who are cheap conditional on their risk score.

This point is empirically important because statutory overpayments have increased substantially over time. From 2001–2003, MA plans would receive 104 percent of FFS costs, absent any risk-selection, as policymakers began to set county “benchmarks” above average county FFS spending.<sup>11</sup> From 2004–2006 (our post-period), these payments rose to 108 percent, both because the county “benchmarks” increased at a faster rate than did FFS spending and because CMS explicitly gave MA plans so-called “budget neutrality” payments (plans argued they would need these extra payments to compensate them for the expected revenue loss due to risk adjustment). From 2007–2009, benchmark increase and budget-neutrality payments led to statutory overpayments between 113 and 114 percent.<sup>12</sup> For this reason, we choose a rather short post-period, when the change in these statutory overpayments is still relatively limited. Consistent with the prediction that once statutory overpayments are sufficiently large plans will be less selective, MA enrollment between 2006 and 2010 increased by 63 percent (from 6.8 million to 11.1 million).<sup>13</sup>

Second, our assumptions regarding the manner in which screening costs vary across risk categories are obviously central to the model, but we are rather silent on what, in practice, these screening costs might entail. How do plans differentially attract “overpriced” consumers? We empirically explore some possibilities in Section VI. Strictly speaking, how they do so is irrelevant to the government’s bottom line, which is our main focus. Of course, it is highly relevant to consumer surplus, which we explore when we discuss welfare in Section VII.

Third, we are also rather silent on plan competition. We believe competition is likely second-order in determining the cost to the government, as MA capitation payments are set by the risk-adjustment formula and not competitive bidding. Again, however, competition likely affects how producer and consumer surplus change as a result of risk adjustment and so we explore this topic in Section VII.<sup>14</sup>

### III. Data

Our empirical work relies chiefly on individual-level data from the Medicare Current Beneficiary Survey (MCBS) Cost and Use series from 1994 to 2006. The MCBS

<sup>11</sup> Throughout the 1980s and early 1990s, policymakers argued that MA plans should be more efficient than FFS and thus systematically set benchmarks at 95 percent of county FFS per capita spending. The earlier work we cited suggests that plans were still able to enroll individuals who were on average less than 95 percent of average FFS costs, meaning net overpayments were still high, but during this period none of the net overpayments were due to statutory overpayments, unlike recent years.

<sup>12</sup> All figures come from various MedPAC reports, which annually document statutory overpayments.

<sup>13</sup> Indeed, while in Section IV we document evidence of significant intensive-margin screening from 2004 to 2006, evidence from more recent years is mixed. In its 2012 annual report to Congress, MedPAC, citing a working-paper version of our study, found nearly identical levels of intensive-margin selection levels using the universe of Medicare enrollees in 2007 and 2008 (MedPAC 2012). Results from McWilliams, Hsu, and Newhouse (2012), however, suggest that risk-selection decreased in 2007–2008 relative to 2004–2006, consistent with higher statutory overpayments diminishing risk-selection incentives.

<sup>14</sup> Past studies have explored consumer surplus, profits, and competition in the MA market, though all are from the pre-risk-adjustment era. Hall (2011) finds that between 1999 and 2002, annual consumer surplus surpassed \$12 billion. Town and Liu (2003) estimate that between 1993 and 2000, the MA program generated over \$18 billion in consumer surplus, and nearly three times that amount in insurer profits.

links CMS administrative data to surveys from a nationally representative sample of roughly 11,000 Medicare enrollees each year. It also provides complete claims data from hospital admissions, physician visits, and all other Medicare-covered provider contact for all FFS enrollees in the sample, totaling about 0.5 million claim-level observations annually. The MCBS follows a subsample of respondents for up to three or four years, thus creating a mix of cross-sectional and panel data. During our sample period, the data comprise more than 55,000 unique individuals and 150,000 person-year observations.<sup>15</sup>

The MCBS records whether an individual is in an MA plan or FFS each month he is in the sample. As noted in Section I, MA plans do not submit claims or costs to CMS, and thus the MCBS only contains claims and health care cost data for those in FFS. Otherwise, all demographic and survey data are recorded for both MA and FFS enrollees. Consistent with past work, we find that, relative to their FFS counterparts, MA enrollees are more likely to live in metro areas, are less likely to be on Medicaid or Social Security Disability Insurance (SSDI) and, conditional on not being on SSDI, are younger.

Some of the key predictions from the theoretical framework in Section II involve enrollees' risk scores, which the MCBS does not report. We obtained risk scores from 2004 to 2006 for all MCBS respondents directly from CMS. However, testing two of our key predictions also involves knowing what individuals' risk scores *would have been* in the earlier years had the HCC formula been in place, which we must generate ourselves. As described above, an individual's risk score in year  $t$  is based on diagnoses documented on claims from year  $t - 1$ . As such, using CMS's algorithm for converting claims data into risk scores, we simulate the risk score for all MA enrollees the year immediately after they switch from FFS. As we know the actual risk scores from 2004 to 2006, we check our simulation in these years: the correlation between our simulated risk scores and CMS's actual scores is more than 0.96.

The need to calculate HCC scores in the pre-period means we limit some of our analysis to those individuals who were in FFS all 12 months of a baseline year, so that we observe their complete claims history that year. Table 2 shows the number of observations who are in FFS in year  $t$  and in MA in year  $t + 1$ , as well as the number who are in FFS both years, and how these numbers change across our sample period. Of the more than 85,000 cases in which we observe a person in both year  $t$  and year  $t + 1$ , more than 1,500 involve switches from FFS to MA. One limitation of the focus on FFS-to-MA switchers is that it ignores those who join MA immediately upon their Medicare eligibility. Our MCBS data demonstrate, however, that more than three-in-four new MA enrollees come from FFS, as opposed to joining when first enrolled in Medicare.<sup>16</sup>

<sup>15</sup>We exclude the 0.25 percent of enrollees whose Medicare eligibility is based on having end-stage-renal disease, as different MA rules apply to them. We also exclude the roughly 2 percent of observations in which the person joins the MCBS in the middle of the year because much of their data are imputed.

<sup>16</sup>In Table 2, we show that 2.1 percent of FFS recipients switch into MA the next year. With an average of 35 million in Medicare FFS during our study period, that represents 735,000 per year. During our period, about 2 million individuals become eligible for Medicare each year, and about 12 percent (240,000) of them are enrolled in MA in that first year.

TABLE 2—FREQUENCY DISTRIBUTION OF TRANSITIONS BETWEEN FFS AND MA, 1994–2006

	Baseline year $t$ equals:				Total
	1994–1996	1997–1999	2000–2002	2003–2005	$t = 1994–2005$
FFS (year $t$ ) → FFS (year $t + 1$ )	19,017	18,539	18,305	17,329	73,190
FFS (year $t$ ) → MA (year $t + 1$ )	566	399	102	464	1,531
MA (year $t$ ) → FFS (year $t + 1$ )	102	165	457	125	849
MA (year $t$ ) → MA (year $t + 1$ )	1,457	3,282	2,805	2,496	10,040
In sample both years	21,142	22,385	21,669	20,414	85,610
Left sample after baseline year	13,883	14,301	14,983	14,284	57,451
Total observations (baseline year)	35,025	36,686	36,652	34,698	143,061

*Note:* An individual in a given year is classified as being on MA if she is on MA for at least half of the months for which she is Medicare eligible in that given year.

#### IV. How Did Selection Patterns into MA Change after Risk Adjustment?

In this section, we empirically test our model's predictions regarding the effect of risk adjustment on both extensive-margin selection (did MA beneficiaries' risk scores rise?) and intensive-margin selection (did their costs conditional on their risk score fall?).

##### A. Quantifying the Selection Incentives Created by the HCC Model

Column 1 of Table 3 presents the average difference between the HCC-based capitation payment and the traditional demographic-based capitation payment using our MCBS data, with this difference broken down by percentiles of the HCC risk score.<sup>17</sup> Mechanically, capitation payments must, on average, rise under the HCC formula for those with higher risk scores, and column 1 merely presents the magnitudes. For example, the HCC capitation payment would, on average, pay about \$3,000 less than the demographic-based capitation payment for individuals with HCC scores in the lowest quartile, but would pay roughly \$7,000 more for the individuals in the top quartile.

Column 1 suggests that insurers would have an incentive to increase risk scores over the entire risk score distribution, but column 2, which reports the HCC capitation payment minus actual costs, shows that doing so would not always be profitable. For example, individuals with the highest 1 percent of risk scores represent, on average, a nearly \$6,000 loss to an MA plan, consistent with the research cited earlier showing that the HCC formula under-predicts costs for enrollees with the most severe disease conditions. Plans might thus be reluctant to draw from the extreme right tail of the risk-score distribution.

Though not shown in the table, we also calculate that the share of individuals who have actual baseline costs less than their risk scores would predict is 77 percent under both the demographic and the HCC model. This result arises because of the

<sup>17</sup> We estimate these payments using pre-2004 data so that selection in reaction to the HCC model has not taken place. After 2003 MA plans enjoyed higher benchmarks as well as additional payments to ease the transition to risk adjustment, which we remove for the purposes of this table. As such, it reflects the change in incentives from "payment-neutral" risk adjustment as defined in Section II.

TABLE 3—SUMMARIZING CHANGES IN INCENTIVES AFTER RISK ADJUSTMENT

HCC score	HCC payment minus demographic payment	HCC payment minus actual Medicare expenditure
0–25th percentile (lowest scores)	–2,993	67
25–50th percentile	–2,406	198
50–75th percentile	–342	549
75–99th percentile	6,701	893
99–100th percentile (highest scores)	29,789	–5,907
Total	491	359
Observations	54,369	54,369

*Notes:* All data taken from the “pre-period” before implementation of risk adjustment, among the subsample of individuals who were in the FFS system all 12 months of the previous year. Both columns use claims data from the previous year to calculate capitation payments under the HCC model for each individual. The first column follows the formula of the demographic model to calculate capitation payments for all individuals. Dollar amounts are adjusted to 2007 dollars using the CPI-U. Sample weights provided by the MCBS are used.

extreme right-skew of health costs—the vast majority of the distribution falls below the mean (or conditional mean, in the case of risk adjustment). Thus, risk adjustment does not actually decrease the number of individuals who are “overpriced”—though it obviously changes their likely characteristics—and indeed we find little change in MA market share after the HCC model is introduced.<sup>18</sup>

### B. Empirical Strategy

Our prediction that “extensive-margin” selection should fall after the shift to risk adjustment would imply a positive estimate for  $\beta$  in the following difference-in-differences specification:

$$(1) \quad \text{Risk score}_{it} = \beta \text{MA}_{it} \times \text{After } 2003_t + \gamma \text{MA}_{it} + \delta_t + \epsilon_{it},$$

where  $i$  indexes the individual,  $t$  the year,  $\text{Risk score}_{it}$  is the individual’s HCC score (which, by definition, uses year  $t - 1$  claims data to predict Medicare expenditure in year  $t$ ),  $\text{MA}_{it}$  the share of her Medicare-eligible months that the individual spends in MA in year  $t$ ,  $\text{After } 2003_t$  the post-period indicator, and  $\delta_t$  a vector of year fixed effects.<sup>19</sup> We estimate this regression on the sample of individuals who are in FFS all 12 months of the baseline year  $t - 1$  so that we can use their complete claims data that year to calculate year  $t$  risk scores.<sup>20</sup>

<sup>18</sup>This analysis does not imply 77 percent of individuals are potentially *profitable*, as there are screening costs and MA might be more or less efficient than FFS in providing the basic Medicare benefits package. Note also that MA enrollments increase substantially after our post-period, in reaction to statutory overpayments averaging over 13 percent.

<sup>19</sup>Both equations (1) and (2) are parsimonious in that they do not control for demographic or other characteristics of the beneficiaries. This choice is deliberate, and it reflects the fact that MA plans are paid based on the risk scores of their beneficiaries, not their risk scores conditional on, say, age.

<sup>20</sup>While we could use the actual risk scores provided by CMS for the post-period, we instead use our simulated risk scores in both the pre- and post-periods so that any change in risk scores will not be driven by differences in how they are calculated. Using actual risk scores in the post-period increases the magnitudes and statistical significance of the coefficients of interest in both the extensive- and intensive-margin analyses.

We next investigate our “intensive-margin” prediction that after risk adjustment, plans will enroll individuals who have low baseline costs conditional on their risk score. This would imply a negative estimate for  $\beta$  in the following specification:

$$(2) \quad \begin{aligned} \text{Expenditure}_{i,t-1} = & \beta MA_{it} \times \text{After } 2003_t + \gamma MA_{it} \\ & + \lambda \text{Risk score}_{it} + \delta_t + \epsilon_{it}, \end{aligned}$$

where  $\text{Expenditure}_{i,t-1}$  is the total FFS expenditure for individual  $i$  in year  $t - 1$  and all other notation and sampling follows that in equation (1).

### C. Results

We begin by exploring how the difference in average baseline Medicare spending changes after risk adjustment among those switching to MA versus those remaining in FFS, and then decompose this effect into its extensive- and intensive-margin components. Column 1 of Table 4 shows that before risk adjustment, those switching to MA have average Medicare spending \$2,847 below those who remain in FFS, consistent with positive selection into MA. The statistically insignificant estimate of  $-\$173$  for the *After 2003* interaction suggests that risk adjustment has little effect on this difference.

The next five columns of Table 4 explore the first component of the decomposition. We report the mean of the dependent variable (roughly 1.1) and online Appendix Figure 2 displays a histogram. Column 2 suggests that while individuals switching into MA before risk adjustment had average risk scores roughly 0.305 points lower than those remaining in FFS, risk scores of those switching into MA rise significantly (by 0.106) after risk adjustment is introduced, making up about one-third of the difference.

Based on the results from Table 3 that outliers in the right-tail are still underpriced by the HCC formula, we expect the effect on the mean to be muted, as plans would still find it unprofitable to enroll those with extreme risk scores. Indeed, in column 3, merely dropping observations with risk scores above the 99th percentile increases the magnitude of the estimate. Estimating a median regression (column 4) on the entire sample increases the coefficient by nearly one-third (to 0.140). While we prefer to use a long pre-period to improve precision by increasing the number of individuals in the pre-period switching from FFS to MA, column 5 shows that excluding observations before 1997 does not change the results. Finally, in column 6, we show the result is robust to controlling for  $MA \times \text{year}$  pre-trends.<sup>21</sup>

Following Bitler, Gelbach, and Hoynes (2006), to get a clearer picture of the effects across the entire risk score distribution, we estimate quantile regressions for the first through 99th quantiles, and plot the resulting coefficients in Figure 2. As predicted (because of the greater variance at higher risk scores), the estimate is generally increasing with the risk score. But it falls close to zero right before the

<sup>21</sup> As risk adjustment phases in between 2004–2006, we see an increasing trend in the post-period. As such, in column 6 we estimate pre-period trends and project them forward to the post-period. The coefficient ( $p$ -value) on the  $MA \times \text{year}$  variable in the pre-period is  $-0.0027$  (0.875), suggesting essentially no pre-trends.



TABLE 4—CHANGES IN SELECTION PATTERNS AFTER RISK ADJUSTMENT

	Cost (1)	Extensive-margin				
		Score (2)	Score (3)	Score (4)	Score (5)	Score (6)
Share of year in MA	−2847.0*** [396.9]	−0.305*** [0.0355]	−0.241*** [0.0350]	−0.233*** [0.000161]	−0.236*** [0.0510]	−0.250*** [0.0350]
Share of year in MA × after 2003	−172.7 [713.4]	0.106* [0.0614]	0.133** [0.0604]	0.140*** [0.000413]	0.129* [0.0709]	0.155** [0.0604]
HCC score (calculated from claims)						
HCC score from CMS						
Mean, dependent var.	6,315.6	1.149	1.091	1.149	1.115	1.091
Estimated method	OLS	OLS	OLS	Q. reg.	OLS	OLS
Outliers trimmed	No	No	Yes	No	Yes	Yes
1998–2006 only	No	No	No	No	Yes	No
Trend control	No	No	No	No	No	Yes
Pre- and post-periods?	Both	Both	Both	Both	Both	Both
Observations	73,054	73,054	72,274	73,054	54,029	72,274
	Intensive-margin					
	Cost (7)	Cost (8)	Cost (9)	Cost (10)		
Share of year in MA	171.5 [316.5]	191.6 [467.2]	106.6 [316.5]			
Share of year in MA × after 2003	−1,217.9** [604.0]	−1,280.3* [691.8]	−1,052.4* [604.0]	−1,656.9*** [516.1]		
HCC score (calculated from claims)	9,903.4*** [182.5]	9691.0*** [202.2]	9,903.4*** [182.5]			
HCC score from CMS				10,653.2*** [452.3]		
Mean, dependent var.	6,315.6	6,482.0	6,315.6	7,372.1		
Estimated method	OLS	OLS	OLS	OLS		
Outliers trimmed	No	No	No	No		
1998–2006 only	No	Yes	No	No		
Trend control	No	No	Yes	No		
Pre- and post-periods?	Both	Both	Both	Post		
Observations	73,054	54,646	73,054	17,680		

Notes: All observations are in FFS all 12 months of the given year. Year fixed effects included in all regressions. The outcome in column 1, and 7 through 10 is an individual's current year total Medicare expenditure. The outcome in columns 2 through 6 is an individual's HCC score the following year, which is based on current-year claims. "Q. reg" refer to median regressions. "Outliers trimmed" excludes individuals with risk scores above the 99th percentile (where percentiles are calculated separately by year). "Trend controls" adjust for  $MA \times year$  based on pre-period trends. In all columns except the last, we calculate HCC scores from the MCBS claims data; the final column uses the HCC score provided directly from CMS (unavailable for the pre-period). Sample weights provided by the MCBS are used. Dollar amounts are adjusted to 2007 dollars using the CPI-U. Standard errors are clustered by the individual.

\*\*\*Significant at the 1 percent level.

\*\*Significant at the 5 percent level.

\*Significant at the 10 percent level.

99th quantile, consistent with outliers being substantially underpriced by the formula, though of course precision is more limited at the highest percentiles.

Given that average Medicare spending of those switching to MA relative to those remaining in FFS does not change after 2003 while their risk scores rise,

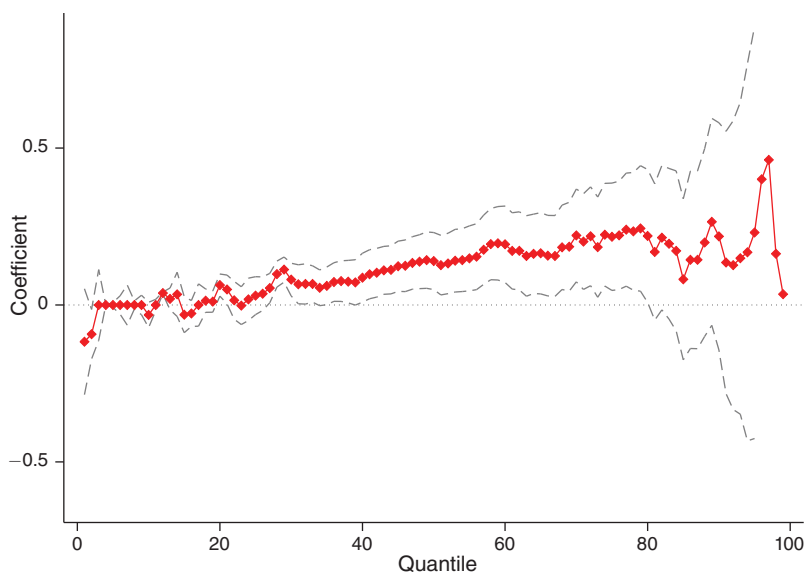


FIGURE 2. COEFFICIENTS AND 90 PERCENT CONFIDENCE INTERVALS  
FROM QUANTILE “EXTENSIVE-MARGIN” REGRESSIONS

Notes: Each point is the coefficient  $\beta$  from a quantile regression of the form  $Risk\ score_{it} = \beta MA_i \times After_t + \gamma MA_i + \delta_t + \epsilon_{it}$ . See Section IV for further details. Note that the confidence-intervals for quantiles greater than the 96th are suppressed so as not to compress the scale of the figure.

intensive-margin selection must have increased. As expected, column 7 shows that, relative to the pre-risk-adjustment period, after 2003 individuals switching into MA versus those remaining in FFS have baseline costs over \$1,200 less than their risk scores would predict. As with the extensive-margin results, the coefficients of interest are robust to excluding years before 1997 (column 8) and controlling for pre-trends (column 9).<sup>22</sup>

In the final specification, we focus exclusively on the 2004 through 2006 period and use the actual risk-scores provided to us by CMS instead of a simulated risk score. Here the intensive margin results are even stronger and more precisely estimated, demonstrating that those joining MA after the shift to risk adjustment have significantly lower costs than their HCC risk scores would predict. This larger effect is expected, as our simulated risk scores (though highly correlated to the official risk scores provided to us by CMS) presumably contain some error and lead to attenuation bias. Finally, note that the main effect of MA status in the intensive-margin regressions is of theoretical interest. The fact that it is close to zero suggests that, among beneficiaries switching to MA in the pre-period, the HCC risk score successfully predicts costs. This result supports the model’s prediction that risk adjustment *would have* reduced selection had the population of individuals joining MA not changed in response to the policy.

<sup>22</sup>Like the extensive-margin analysis,  $MA \times year$  trends in the pre-period are essentially zero: a coefficient of  $-20.69$  with a  $p$ -value of  $0.925$ .

TABLE 5—CHANGES IN TOTAL BASELINE MEDICARE EXPENDITURE  
FOR MA VERSUS FFS ENROLLEES AFTER RISK ADJUSTMENT

	Costs (1)	Charges (2)	Part A (3)	$\Delta \ln(\text{FFS cost})$ (4)
Share of year in MA	−2,847.0*** [396.9]		−1,279.5 [882.3]	
Share of year in MA × after 2003	−172.7 [713.4]		−235.9 [997.5]	
In MA at discharge		−1,497.5*** [278.4]		
In MA at discharge × after 2003		85.50 [409.2]		
$\Delta \ln(\text{FFS share})$				−711.9** [329.0]
$\Delta \ln(\text{FFS share})$ × after				−86.06 [697.0]
Mean, dependent var.	6,315.6	25,559.6	3,489.3	298.4
Dataset used	MCBS	Hosp.	MCBS	County FFS
Years	1995–2006	2000–2006	2000–2006	2000–2006
Observations	73,054	8,217,647	42,051	18,658

Notes: Column 1 replicates column 1 of Table 4. Column 2 uses hospital discharge data and estimates total charges as a function of MA status, hospital, and year fixed effects. Column 3 replicates column 1 of Table 4 but uses only part A costs and years from 2000–2006, to be more comparable to the hospital discharge analysis. Column 4 uses county-level data and replicates the analysis in Batata (2004). She shows that in the following regression  $\Delta \ln(\text{Avg. FFS costs})_{ct} = \alpha + \beta \Delta \text{FFS share}_{ct} + \epsilon_{ct}$ , where  $\text{Avg. FFS costs}_{ct}$  is the average per capita costs for all FFS enrollees in county  $c$  in year  $t$  and  $\text{FFS share}_{ct}$  is the share of county  $c$ 's Medicare enrollees in FFS in year  $t$ ,  $\beta$  is an estimate for the difference between the costs of the *marginal* enrollee switching between MA and FFS and the *average* FFS enrollee. All regressions are estimated via OLS and for regressions using MCBS data the provided weights are used.

\*\*\*Significant at the 1 percent level.

\*\*Significant at the 5 percent level.

\*Significant at the 10 percent level.

#### D. Discussion and Further Verification

One drawback of our identification strategy is that, because we need to calculate risk scores in the pre-period, we focus on individuals who are in FFS in a baseline year and identify our coefficients off of those who switch the next year from FFS to MA.<sup>23</sup> As noted earlier, CMS provided us with actual risk scores for all individuals in the MCBS from 2004 to 2006. While we cannot replicate the specification in Table 4, we can determine whether—consistent with our findings for MA switchers—risk scores among all MA enrollees are growing faster than those in FFS during this period. We find that in the CMS administrative data, the average MA risk score increased by 12 percent from 2004 to 2006, versus just 1 percent for those in FFS. As such, our estimates above comparing FFS-to-MA “switchers” versus FFS

<sup>23</sup>In the same report that we cited in footnote 13, MedPAC also examined MA-to-FFS switchers from 2007–2008, whereas this group is too small in the MCBS for us to examine. Consistent with our model, they find that MA enrollees who switch back to FFS are *expensive for their risk score*, the parallel result to finding that FFS-to-MA switchers are *cheap for their risk score*.

“stayers” closely correspond to comparisons using the stock of individuals in MA versus FFS.

In Table 5, we move beyond the MCBS to verify our finding of no overall cost differences between FFS and MA enrollees after risk adjustment, as we found in column 1 of Table 4. We gathered over 8 million hospital discharge records from the 12 states between 2000 and 2006 that require hospitals to record FFS/MA status. Because these states are populous, they represent 42 percent of Medicare beneficiaries during this period. In column 2 of Table 5, the average MA patient had \$1,500 less in charges than his FFS counterpart before 2004—consistent with positive selection in the pre-period—with only \$85 ( $p = 0.835$ ) of this difference being made up in the post-period—consistent with our results that risk adjustment did not lead MA plans to enroll higher-cost individuals. Interestingly, when we do our best to replicate this regression with the MCBS switcher-analysis by using only annual part A charges in 2000 to 2006, we find very similar point-estimates (column 3).

Finally, we follow Batata (2004) and use county-level data to estimate changes in MA selection. She shows that regressing county-level changes in average FFS spending on the change in county-level MA penetration yields a measure of the difference in costs between the *marginal* person switching between MA and FFS and the FFS stock. While slightly different than our switcher regressions—which compare the *average* person switching from FFS to MA with the *average* person staying in FFS—one would expect these two selection measures to move in the same direction. As the final column of Table 5 shows, this difference is negative before risk adjustment, reflecting the fact that those on the margin of switching between MA and FFS have lower costs than those in FFS. Consistent with the switcher analysis and the rest of the table, selection along this margin does not change after risk adjustment (in fact, the point estimate suggests increased selection), though our precision here is somewhat limited.

In summary, our evidence from several datasets indicates that, with respect to actual health costs, those in MA are as positively selected after risk adjustment as before.

## V. Did Risk Adjustment Decrease Differential Selection?

One might assess risk adjustment by estimating how an individual's annual *Total Medicare expenditure* changes when he switches from FFS to MA, and then compare this change before and after risk adjustment. *Total Medicare expenditure* is the total annual cost to Medicare for covering an individual, whether from claims (for FFS enrollees) or capitation payments (for MA enrollees). Under perfect risk adjustment (i.e., capitation payments equal to an individual's expected FFS costs) whether an enrollee switches between FFS and MA should in expectation have no effect on this variable.

This approach has important limitations. First, comparing the government's costs as individuals switch between FFS and MA obviously requires focusing only on “switchers,” and thus only a subsample of the data. Yet our model tells us that risk adjustment will decrease overpayments on the *stock* of those who joined MA before risk adjustment, while having ambiguous effects on those who join after, meaning that looking only at post-period “switchers” could mask the ability of risk

adjustment to decrease overpayments. On the other hand, using risk scores only from those who were in FFS the previous year ignores “intensive coding” (as risk scores that first year are based on FFS claims), a serious drawback of risk adjustment that a “switcher” analysis cannot measure.

Second, after 2003, the MCBS does not provide individual capitation payments. In principle, one can recreate them by multiplying the simulated HCC risk scores by county-level benchmarks. However, recall that county benchmarks grow more rapidly post risk adjustment and plans also receive additional “budget-neutrality” adjustments to ease the transition to risk adjustment. As such, one must take a stand on the counterfactual evolution of benchmark payments to isolate the effects of risk adjustment from these coincident payment increases.

In online Appendix B, we make assumptions about how each of these missing pieces affects our calculation and conclude that overpayments do not fall post-risk adjustment. We in fact find a small positive effect—a significant increase in overpayments among those who enroll in MA post-2003 of between \$1,500 and \$2,000, somewhat but not fully offset by a \$700 decrease in overpayments among the still larger population of MA “incumbents” who had first joined before 2004—though given the above concerns large error bands must be assumed. Note also that this blended effect should become more positive over time as the incumbents comprise a shrinking share of the MA population given its substantial flux.

In the rest of the section, we focus on a specification that allows us to examine the change in differential selection for *all* MA and FFS enrollees—not just switchers—before and after risk adjustment, in a manner independent of changes in underlying benchmarks. Specifically, we use mortality as a proxy for costs—which, unlike costs, is both recorded for everyone in the MCBS and independent of how the government changes benchmarks—and regress it on MA status and the risk score. A negative coefficient on the MA variable indicates that, even conditional on the risk score, MA enrollees are positively selected. We compare whether this selection conditional-on-the-risk-score is greater in the pre-period (using the demographic risk score) or the post-period (using the HCC risk score).<sup>24</sup>

Below, we show that our demographic and HCC scores predict the correct amount of variance in the FFS data; that mortality is indeed an excellent proxy for costs in FFS data; and then proceed to the main empirical test.

### A. Empirical Results

*Initial Steps.*—As noted in Section I, the demographic risk score was shown by CMS to account for 1 percent of FFS cost variance. Column 1 of online Appendix Table 4 shows that the demographic risk scores we calculate account for 1.23 percent of cost variance, using pre-period FFS data. Similarly, CMS calculated that the HCC score accounts for 11 percent of pre-period FFS cost variance. Indeed, in column 2 our  $R^2$  value using MCBS FFS data is also 11 percent. Column 3 regresses annual costs on whether an MCBS respondent died in a given year. The  $R^2$  value is

<sup>24</sup>From CMS, we have HCC risk scores for everyone (both switchers and stayers) in the post-period, and the demographic risk scores can be easily calculated with the background information collected by the MCBS and does not require claims data.

TABLE 6—CHANGES IN DIFFERENTIAL SELECTION BEFORE AND AFTER RISK ADJUSTMENT

	Died within the calendar year ( $\times 100$ )						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of months in MA	-1.414*** [0.158]	-1.440*** [0.182]	-1.345*** [0.319]	-0.310* [0.181]	-0.923*** [0.312]	-0.310* [0.181]	-0.253 [0.231]
Share of months in MA $\times$ after 2003						-0.613* [0.361]	-0.669* [0.389]
Demographic score				4.675*** [0.0906]			
HCC score					5.563*** [0.226]		
Demogr. score $\times$ pre-period						4.675*** [0.0906]	4.659*** [0.123]
HCC score $\times$ post-period						5.563*** [0.226]	5.563*** [0.226]
Mean, dependent var.	0.0497	0.0495	0.0502	0.0495	0.0502	0.0497	0.0495
Period	Both	Pre	Post	Pre	Post	Both	Both
Short pre-period	No	No	No	No	No	No	Yes
Observations	137,769	105,364	32,405	105,364	32,405	137,769	91,802

Notes: Observations are no longer restricted to being in FFS the previous year. The demographic risk score is calculated for all MCBS observations using the formula provided by CMS. The HCC risk score is provided directly by CMS for all MCBS respondents after 2003. Year effects included in all regressions. Sample weights provided by the MCBS are used. Standard errors are clustered by the individual.

\*\*\*Significant at the 1 percent level.

\*\*Significant at the 5 percent level.

\*Significant at the 10 percent level.

15.1 percent, showing that death is in fact a stronger predictor of FFS costs than the HCC risk score itself.

*Comparing Selection Before and After Risk Adjustment.*—The first three columns of Table 6 show that—unconditional on the risk score—those in MA are very positively selected with respect to mortality. They are 1.4 percentage points (more than 25 percent) less likely to die in a given year, relative to their FFS counterparts, which holds relatively constant before and after risk adjustment (columns 2 and 3). The nearly identical mortality advantage of those in MA pre- and post-risk adjustment adds to the evidence shown in Table 5 that the differences in overall, unconditional health status between MA and FFS enrollees do not change after risk adjustment.

To test risk adjustment, however, we compare *conditional* differences. In the pre-period, the demographic risk score was more effective in reducing this positive selection—the MA coefficient, while still negative and significant, falls in magnitude by roughly 75 percent in column 4 relative to column 2. In the post-period, conditioning on the HCC risk score in column 5 only modestly reduces the positive selection into MA—the coefficient on the MA variable falls by less than 30 percent and remains highly significant.

In column 6, we combine the regressions in columns 4 and 5 so that we can more easily compare the MA coefficients in the pre- and post-periods. Indeed, the coefficient on MA in the post-period is significantly “more negative” than that in the pre-period, suggesting that *conditional on the risk score that MA plans faced at the*



*time*, being in MA indicates a lower conditional probability of dying and thus more positive selection. The final columns show that this relationship holds when we compare the post-period to a shorter pre-period.

### B. Discussion

We can translate these effects into overpayments by using the estimated effect of mortality on total annual Medicare costs. This exercise yields an increase in overpayments of \$317. An advantage of the mortality analysis is that it provides an estimate of differential payments independent of the extra statutory overpayments MA plans received, but as a policy matter, MA plans did indeed enjoy higher statutory overpayments in the post-period. Including them yields total overpayment increases between \$736 and \$988, the smaller estimate equal to roughly 9 percent of average per capita FFS annual spending.<sup>25</sup>

## VI. How Does Selection into MA Plans Take Place?

### A. Why are Low-Cost Individuals More Likely to be in MA Plans?

The evidence in Section IV shows that MA plans enrolled lower-cost individuals both before and after risk adjustment. But how do such patterns emerge when plans must offer the same plans at the same rate to all Medicare beneficiaries in their geographical area of operation? We first explore whether among all MA enrollees, the healthy ones are more satisfied with their care and less likely to return to FFS. This pattern might arise because plans actively treat healthy enrollees better than sick ones so as to differentially retain the former group, or simply because sick individuals do not like the HMO model of care. Through reputation effects, such a result could feed back into patterns of switching into MA as well. This latter pattern could also be driven by targeted advising, with previous studies finding that advertisements for MA plans target healthy people (Mehrotra et al. 2006; KFF 2008).

The MCBS asks respondents to rate their satisfaction with their overall health care “last year” as well as specific aspects of it. As the question is asked in the fall, it is difficult to know whether individuals are answering based on their experience so far in the current year or in the previous calendar year. As such and in contrast to our previous analyses, here we focus on those who did not switch (either from FFS to MA or from MA to FFS) the previous year by comparing individuals in MA in both years with those in FFS in both years.

<sup>25</sup> The monthly increase in costs the year an individual dies is \$4,308 (online Appendix Table 4, column 3). The increase in conditional selection with respect to mortality after risk adjustment is 0.00613 (Table 6, column 6), suggesting an increase in overpayments of  $\$4,308 \times 12 \times 0.00613 = \$317$  (assuming the relationship between costs and mortality is the same for MA and FFS). The increase in statutory overpayments depends on which pre-period to use as a counterfactual. As noted in Section II, in the three years before the reform, statutory overpayments averaged 103 percent of FFS costs, rising to 108 percent in our post-period. Taking the entire pre-period, where in the early years capitation payments were set to equal 95 percent of per capita FFS costs, we can estimate that capitation payments were roughly 100 percent of FFS costs. Given that per capita FFS costs were \$8,385 in 2004, the statutory overpayment increases are between  $0.05 \times \$8,385 = \$419$  and  $0.08 \times \$8,385 = \$671$ , so, adding the increase in overpayments from differential selection alone gives a total overpayment increase between \$736 and \$988.

Ideally, we would explore whether satisfaction is higher among MA recipients who are more profitable to insurers. But because we do not have health care cost or claims data for those in MA, we use self-reported health as an admittedly imperfect proxy and investigate whether good health predicts satisfaction with one's health care in MA more than in FFS:

$$(3) \quad \text{Satisfaction}_{it} = \beta \text{MA}_{it} \times \text{Health}_{it} + \gamma \text{MA}_{it} + \mathbf{H}_{it} + \lambda X_i + \delta_t + \epsilon_{it}.$$

In this specification, *Satisfaction* measures individuals' reported satisfaction with different aspects of their health care and varies from one (very dissatisfied) to four (very satisfied), *Health* is a five-category self-reported health variable,  $\mathbf{H}$  are its corresponding fixed effects, and all other notation follows that used in previous equations. The health fixed effects account for the fact that in both MA and FFS, poor health correlates with negative feelings toward one's health care. Thus the interaction term explores how much more or less sensitive enrollee satisfaction is to underlying health in MA versus FFS. We control for demographic characteristics in  $X$  because different groups may assess their health and health care differently. If MA plans treat healthier enrollees better, we would expect  $\beta > 0$ .

Table 7 displays the results from estimating equation (3) via OLS. We demean the *Health* variable in  $\text{MA} \times \text{Health}$ , so that the *MA* main effect represents the association with MA enrollment for someone with mean self-reported health. The first row reports results when overall satisfaction serves as the dependent variable. The *MA* main effect is negative—suggesting that someone of average health reports lower satisfaction in MA than in FFS. This estimate is surprising given that MA enrollees self-selected into MA and given the large overpayments to MA plans, which could lead to additional benefits. But then again MA enrollees may simply be harder to please.

We instead focus on the coefficient on the interaction term, which is positive and significant, indicating that good health predicts satisfaction with MA plans more than it does satisfaction with FFS. In fact, only among those who report being in “excellent” health do MA plans receive a higher rating than FFS (not shown). Moreover, relative to FFS enrollees, MA enrollees exhibit a more positive gradient of satisfaction with respect to health in all nine categories surveyed by the MCBS, and for a majority of categories this difference is statistically significant.

The last row of Table 7 investigates whether sicker MA enrollees “vote with their feet” and exit at higher rates than do sicker enrollees in FFS. Instead of satisfaction ratings, we regress whether an individual changes his coverage status—to MA if he is currently in FFS, to FFS if he is in MA—on the same set of explanatory variables.<sup>26</sup> Indeed, the same pattern emerges—not only are MA enrollees less likely to retain their current coverage status in general, but this difference is especially pronounced for those in self-reported poor health. As such, among MA patients, the sickest are the most likely to return to FFS each year.

<sup>26</sup>For this analysis, we expand our MCBS sample to include individuals who switch from FFS to MA (or vice-versa) in each year.

TABLE 7—EFFECT OF MA ENROLLMENT AND HEALTH STATUS ON ENROLLEE SATISFACTION

Dependent variable: Satisfaction rating (1–4)	Mean	Observations	Coefficient (SE)	
			In MA	MA × Health
Overall medical care	3.26	75,884	−0.0225*** (0.00776)	0.0151** (0.00654)
Out-of-pocket costs	3.01	75,309	0.0383*** (0.00912)	0.0226*** (0.00740)
Follow-up care	3.16	69,764	−0.00524 (0.00702)	0.00605 (0.00600)
Doctor's concern for your health	3.15	74,711	−0.0216*** (0.00748)	0.0156** (0.00631)
Information about your medical condition	3.12	75,539	−0.0117* (0.00704)	0.0131** (0.00591)
Access to specialists	3.17	57,187	−0.0343*** (0.00773)	0.000745 (0.00644)
Questions answered over phone	3.06	48,616	−0.0239** (0.00930)	0.0193** (0.00770)
Availability of care nights and weekends	3.11	44,502	−0.00249 (0.00919)	0.00847 (0.00797)
Medicare care provided in same location	3.10	69,380	0.0247*** (0.00698)	0.00890 (0.00573)
Retains coverage type next year	0.97	84,158	−0.0505*** (0.00350)	0.00876*** (0.00284)

Notes: Each row represents a regression of the form:  $satisfaction\ category_i = \beta_1 MA_i + \beta_2 MA_i \times Health_i + \gamma H_i + \lambda X_i + \epsilon_i$ , where *satisfaction* takes values from one to four (“very dissatisfied,” “dissatisfied,” “satisfied,” “very satisfied”), *MA* is a dummy variable for being enrolled in Medicare Advantage at least half of all Medicare-eligible months in a given year, *Health* is a (demeaned) linear measure of the five-category self-reported health variable, *H* is a vector of fixed effect for the five health categories (one, “poor,” up to five, “excellent”), and *X* is a vector of basic controls: age, state-of-residence, year, female, race, disabled, education, income and Medicaid status. As the *Health* variable is demeaned, the coefficient on the *MA* indicator variable represents the effect of being enrolled in MA for an enrollee with average health. A positive coefficient on *MA* × *Health* indicates that the relationship between satisfaction and health status for MA enrollees is greater (“more positive”) than that for FFS enrollees. Note that the sample size varies across regressions because not all questions are asked each year and there is variation in the number of individuals who respond that they do not have enough information to answer. Except for the last outcome, these regressions include only people with the same MA status in both the baseline year and the previous year. Sample weights provided by the MCBS are used and standard errors are clustered by individual.

\*\*\*Significant at the 1 percent level.

\*\*Significant at the 5 percent level.

\*Significant at the 10 percent level.

### B. Possible Mechanisms Underlying the Changes After Risk Adjustment

The results in Table 7 provide an explanation for why higher-cost enrollees tend to be in FFS, but they do not explain how individuals with low costs relative to their risk score found their way into MA plans after risk adjustment. While we cannot provide a definitive answer given the available data, we believe several factors are at work. First, insurers have a wealth of data, both on their own MA enrollees and from their operations in the non-Medicare market. In fact, because insurers (unlike CMS) have data on *the medical claims and costs of MA beneficiaries*, along this dimension plans have more information than the government.

Second, CMS does not adjust for factors such as race, ethnicity, and income, which are not only related to health costs but, through targeted advertising, are also relatively easy for MA plans to select on. Given that the variance in costs grows with the risk score, demographic differences that are small on average could be very large

for groups with high risk scores or for a specific disease category. Plans have the data to determine that, for example, Hispanics with heart disease are \$3,500 cheaper than their risk score would suggest and then target advertising accordingly. If demographic or other observable factors explain how costs vary from the risk score's prediction within a disease category, then plans may have the ability to differentially enroll people who are low cost for their risk score.<sup>27</sup>

We explore empirically whether after risk adjustment MA plans appear to engage in selection along profitable *Disease*  $\times$  *Demographic Group* margins. We start with a specific example. Consider HCC10, "Breast, Prostate, and Colorectal Cancer," which, unlike the other major categories (the ten most common of which are listed in online Appendix Table 1) combines multiple diseases. Given the prevalence rates of these diseases, the large majority of women (men) in this category will have breast (prostate) cancer. Moreover, past medical research (Yabroff et al. 2008) has shown the annual cost of breast cancer treatment to be roughly \$1,900 cheaper than that of prostate cancer even after accounting for demographic differences between the two groups.<sup>28</sup> Therefore, we have a common disease group in which we can identify a large subgroup of individuals (women) who were underpriced before risk adjustment and are overpriced after. While all individuals in Category 10 were less likely than other Medicare beneficiaries to join MA in the pre-period, women—but not men—in this category are more likely to join after risk adjustment ( $p < 0.02$ ).

Next, we explore whether, more generally, MA enrollment in the post-period appears to correspond to *Disease Category*  $\times$  *Gender* "errors" in the risk-adjustment formula. Because of power constraints, we consider only the largest ten disease categories and use gender because other categories create unequal splits of disease groups and thus very small cells. While HCC10 provides an especially nice example given that two diseases are combined, it is also the case that men and women appear to differ systematically in the costs they incur for the other nine disease groups. For each of the 20 cells, we estimate the "error" in the risk-adjustment formula using pre-period FFS data and pre-period benchmarks.<sup>29</sup> As online Appendix Figure 3.A shows, while the errors are roughly centered on zero, there exists substantial variance. On the y-axis, we plot the probability an individual in each cell switches from FFS to MA in the post-period. Indeed, there is a positive correlation between overpayment errors and the post-period FFS-to-MA transition probability. Online Appendix Figure 3.B shows that, by contrast, there is no such correlation between the payment errors and pre-period transition probability—we would expect none, since risk adjustment did not exist in this period and thus risk-adjustment errors would be meaningless to plans.

These results provide additional evidence that insurers responded to the policy-induced change in financial incentives by enrolling those Medicare recipients whose profitability increased most after the shift to risk adjustment.

<sup>27</sup> See recent work by Kim and Aizawa (2013) using advertising spending by MA plans to document that such spending is higher in localities where the cost differences between the sick and healthy are largest and thus the benefits of risk-selection are greatest.

<sup>28</sup> In our data, the gender difference in costs among those with HCC 10 is roughly \$1,200, likely because of attenuation bias from colorectal cancer.

<sup>29</sup> To capture as closely as possible the errors MA plans would encounter, we use data from 1997 to 2002, as the risk-adjustment formula was calibrated using 1999–2000 FFS data and we add two years on either side to gain precision.

## VII. Welfare and Policy Implications

While we have provided evidence suggesting that risk adjustment did not accomplish its goal of reducing the government's overpayments, a full welfare analysis would include the effects on producer and consumer surplus. While the evidence presented in this section is hardly definitive, it begins to shed light on the policy's wider implications.

### *A. The Effect of Risk Adjustment on Plan Profits*

We do not have access to actual MA-specific profit data from insurers, a limitation that appears to be shared by all papers in the MA literature. However, our model does speak directly to plan profits—after risk adjustment, profits fall (even if overpayments rise) because screening costs increase. Note also this prediction refers to “payment-neutral” risk adjustment—which, recall, was our term for a risk-adjustment model in which, essentially, the government is not trying to systematically under- or overpay plans on average.

While we do not have actual profit data to verify the model's prediction, evidence from plans' reactions to risk adjustment suggest they believed payment-neutral risk adjustment would hurt their profits. The implementation of the mild precursor to the HCC model (which, recall, explained 1.5 percent of cost variation) was explicitly payment-neutral. Many private insurers formally called on CMS—which tacitly agreed the reform would hurt plans—to delay the implementation of risk adjustment or to provide extra “budget-neutrality” payments to compensate for risk adjustment.<sup>30</sup> Indeed, perhaps to avoid a similar backlash, as we explain in Sections I and II, CMS increased the risk-adjusted capitation payments by roughly 10 percent as the HCC risk-adjustment formula was phased in. Thus, when risk adjustment is not augmented with additional payments, both CMS and the private insurers expected profits to fall, consistent with our model's predictions.

In our model, risk adjustment can increase government spending while decreasing plan profits because it increases insurers' screening costs. The model assumes that strategies that allow insurers to find, for example, the cheapest diabetics (an optimal strategy after risk adjustment) are costlier on a per-enrollee basis than avoiding diabetics altogether (optimal before risk adjustment). This extra money could be spent in ways that increase consumer welfare, like improving the quality of medical care, or in ways that likely do not, such as engaging in targeted advertising or devising complicated screening strategies. Ultimately, the extent to which this extra spending benefits consumers is an empirical question.

### *B. The Effect of Risk Adjustment on Consumer Welfare*

In this subsection, we explore whether risk adjustment improved consumer welfare along a variety of observable dimensions. To conserve space we only briefly describe the data and estimating equations and refer readers to the online Appendix

<sup>30</sup> See <http://www.cms.gov/MedicareAdvtgSpecRateStats/Downloads/Announcement2000.pdf>. This same document shows that the PIP-DCG model was meant to be explicitly payment neutral.

for greater detail. It is worth emphasizing here that, through the accelerated growth of county benchmarks and “budget neutrality” payments, MA plans received increases in capitation payments in the post-period beyond those that we attribute to their endogenous reaction to risk adjustment. As such, if some of these overpayments were passed on to MA enrollees after 2003, our estimates serve as an upper bound for benefits to consumers from risk adjustment alone.

*Comparing MA versus FFS Before and After Risk Adjustment.*—We begin by examining the simple question of whether MA enrollees report higher satisfaction after risk adjustment. We regress each of our satisfaction measures on  $MA \times \text{After } 2003$ , its lower-order terms and the controls in Table 7. As shown in online Appendix Table 5, the coefficient on  $MA \times \text{After } 2003$  is negative for five and positive for four, though rarely statistically significant, suggesting that post-risk-adjustment, MA enrollees do not report relative increases in their satisfaction.

We probe further on these results by controlling for differential trends among MA recipients prior to the shift to risk adjustment (panel B). In contrast to our previous results for extensive and intensive margin selection, which were unaffected by the inclusion of these trends, here the point estimates generally increase and suggest that MA recipients experienced an increase in satisfaction following the shift to risk adjustment. The increase in the estimated impacts is driven by a negative trend in satisfaction with MA prior to 2004, which is somewhat arrested after 2003. However, in the post-period, MA recipients still report lower satisfaction in eight of the nine categories the MCBS collects (not shown). As such, while there might be some improvement relative to trend, it is not large enough to close the MA-FFS satisfaction gap despite the increase in payments to MA plans in the post-period.

Nor did the positive  $MA \times \text{Health}$  satisfaction gradient documented in Table 7 shift post-risk-adjustment, as one would have predicted had MA plans began to cater to those in worse health. In online Appendix Table 6 we interact  $MA \times \text{Health}$  with a post-risk-adjustment indicator variable, and the only significant effect is a further positive tilt in one category and no statistically significant effect on the others.

*Did Risk Adjustment Improve the Medicare Program More Broadly?*—To fully assess risk adjustment we need to consider its effects on the entire Medicare program. Suppose, for example, that risk adjustment allowed for better sorting of beneficiaries between MA and FFS, and thus everyone was better off. The above analysis would obscure such an effect by not considering the benefits to FFS enrollees as well. Moreover, because the types of people switching from FFS to MA change after 2003 (*e.g.*, they have higher risk scores), the analysis above—comparing MA versus FFS enrollees after 2003—could be contaminated by compositional changes. We thus perform a number of analyses not subject to these critiques.

As envisioned by policymakers, risk adjustment would increase insurers’ incentives to expand health care options for those in poor health. As such, one would expect those in poor health to be relatively more satisfied with their health care options post-risk adjustment. We thus estimate our various satisfaction measures as a function of  $\text{Health} \times \text{After } 2003$ , its lower-order terms, and the controls included in Table 7. As shown in online Appendix Table 7, the coefficient on the interaction



term is positive for all nine categories and more often than not significant, meaning that after risk adjustment those in poor health are relatively *less* satisfied with their health care.

This result is central to the welfare effects of risk adjustment because the gradient of satisfaction with respect to health status speaks to the insurance value of Medicare. Put differently, a system in which healthy people receive the highest quality health care would not seem to allocate resources from the “good” to the “bad” state, as consumption-smoothing requires. The positive coefficients on *Health*  $\times$  *After* suggest this allocation of resources between the good and bad states was potentially made worse after risk adjustment.

We next turn to the National Health Interview Survey to explore whether measures of Medicare enrollees’ health care quality improved after risk adjustment relative to similar individuals outside the program. Using data from 2000 to 2006, we compare the “young elderly” (65–74 year-olds) to the “near elderly” (55–64 year-olds). We examine all variables in the dataset related to patient satisfaction and preventive care. Online Appendix Table 8 shows that none of these measures improve for the young-elderly relative to the near-elderly, as one would have expected if risk adjustment improved care generally for Medicare enrollees.

Finally, we investigate whether MA-intensive counties saw mortality improvements for the young-elderly relative to the near-elderly after 2003. All else equal, in counties with greater MA penetration initially, risk adjustment represents a larger intervention to the Medicare program. We obtained county-level data on mortality by age from the National Center for Health Statistics for 2000–2006. As online Appendix Table 9 shows, we find no evidence of differential mortality improvements among the young elderly in counties most affected by risk adjustment—the coefficients vary in sign but are essentially zero. Nor do changes in MA penetration after risk adjustment predict greater mortality gains for the young-elderly relative to the near-elderly than they did before risk adjustment. We also examine “long differences” to allow the effect a few years to materialize, and if anything, after 2003, increases in MA penetration are associated with a small increase in mortality for the young-elderly relative to the near-elderly, though these results are not quite statistically significant.

In summary, the results from this and the previous subsection suggest little improvement in the care of those in MA versus FFS, nor improvements in the care of Medicare enrollees more generally after risk adjustment.<sup>31</sup>

### C. Improving Risk Adjustment

Our results suggest at least two concerns about future attempts to improve risk adjustment. The first relates to the methodology for recalibration of the risk-adjustment formula, which, essentially, just reruns the cost regressions on the HCC dummies using more recent FFS data. Our results in Table 4 on intensive-margin selection indicate that, after risk adjustment, the lowest-cost cases of each disease category

<sup>31</sup> Our result adds to recent work showing that increases in plan reimbursement have not led to substantial increases in consumer surplus. Cabral, Geruso, and Mahoney (2014) find that about 12 cents for every dollar of MA overpayment increase is passed on to consumers, while Duggan, Starc, and Vabson (2014) find that less than a third are passed on.

become more likely to exit FFS and join MA. As such, using more recent FFS data will create even greater positive bias for a disease category, as the FFS enrollees with the disease are more adversely selected. Put differently, if a disease condition was overpriced initially, recalibration will *exacerbate* the mis-pricing.

The second note of caution is a more general reminder that we believe bears further emphasis: the  $R^2$  of a risk-adjustment formula is not a sufficient statistic for its welfare effects (or even its effect on government payments). As noted, the model shows that the effect of an increase in  $R^2$  on differential payments is indeed ambiguous. We thus add to the argument first made by Glazer and McGuire (2000) that the risk adjustor needs to consider not a formula's "predictiveness," but also the incentives it creates.

Our results do suggest some potential improvements. First, our framework can help predict which disease categories might be especially vulnerable to selection. The results in Section VI suggest that disease categories that have greater predictable variance along dimensions plans could easily observe and target (e.g., gender, ethnicity) will be especially problematic. The risk adjustor may wish to pay special attention to selection in these categories and could consider, for example, conditioning on the variables that empirically predict an individual's cost. In particular, our results suggest including interactions between demographics and certain diseases could be beneficial.

Alternatively, instead of trying to improve the predictiveness of the formula using FFS data, CMS may wish to incorporate the information embodied in MA enrollment patterns into the formula. If a certain disease group begins to "differentially disappear" from FFS and "re-appear" in MA after risk adjustment, then it is a signal that low-cost individuals in that disease group are, for whatever reason, easy to "skim," and thus CMS could reduce the capitation payments for that disease category. Note that such an approach would have the exact opposite effect of the current recalibration procedure, which leads to payment increases for disease groups that experience differential migration to MA.

### VIII. Conclusion

We developed a simple model for understanding how risk selection would respond to an attempt to decrease differential payments to MA plans via risk adjustment. We predicted that MA plans would enroll more Medicare recipients with conditions included in the formula ("extensive-margin" selection falls), but would increase efforts to enroll those with low costs conditional on the risk score ("intensive-margin" selection rises). Using individual-level data on Medicare expenditures and comparing the selection patterns for those switching to MA with those remaining in FFS, we confirmed both predictions. Our framework also shows that because the variance of medical costs increases with the risk score, risk adjustment can increase the scope for selecting individuals with costs below their capitation payment. Indeed, we find that changes in selection led to a small increase in overpayments after risk adjustment and, combined with increases in across-the-board average payments to MA plans, actual overpayments to MA plans increase meaningfully after risk adjustment. We find little evidence that these overpayment increases improved enrollees' quality of care or satisfaction.

The MA program and Medicare more generally have recently been the target of regulation and reform, and as data become available, future work might examine how MA selection and differential payments change in response. The Affordable Care Act (ACA) directly affects the MA program by lowering many county benchmarks while at the same time linking capitation payments to measures of plan quality.

The ACA also requires risk adjustment in the federal and state-run insurance exchanges. While Medicare has access to claims and cost data for its FFS recipients to calibrate a risk adjustment model, no such “public option” will exist in the exchanges, and thus there may be insufficient data to estimate a model.<sup>32</sup> The lack of a public option is important for another reason. Recall from Section VI that both before and after risk adjustment, MA enrollees in poor health express greater dissatisfaction with their care than do their counterparts in FFS, and differentially migrate back to FFS. Plans in the insurance exchanges would seem to face similar incentives not to retain underpriced enrollees and thus might devote limited resources to their care. But these enrollees would not have a public, FFS-like plan to which to return. Thus, while the cost of imperfect pricing in the MA context is primarily borne by taxpayers via higher Medicare spending, we speculate that the underpriced enrollees themselves may bear more of this cost in the exchanges or other settings without a public option.

## REFERENCES

- Batata, Amber.** 2004. “The Effect of HMOs on Fee-For-Service Health Care Expenditures: Evidence from Medicare Revisited.” *Journal of Health Economics* 23 (5): 951–63.
- Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes.** 2006. “What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments.” *American Economic Review* 96 (4): 988–1012.
- Brown, Jason, Mark Duggan, Ilyana Kuziemko, and William Woolston.** 2014. “How Does Risk Selection Respond to Risk Adjustment? New Evidence from the Medicare Advantage Program: Dataset.” *American Economic Review*. <http://dx.doi.org/10.1257/aer.104.10.3335>.
- Bundorf, M. Kate, Jonathan Levin, and Neale Mahoney.** 2012. “Pricing and Welfare in Health Plan Choice.” *American Economic Review* 102 (7): 3214–48.
- Cabral, Marika, Michael Geruso, and Neale Mahoney.** 2014. “Does Privatized Medicare Benefit Patients or Producers? Evidence from the Benefits Improvement and Protection Act.” Unpublished.
- Congressional Budget Office.** 2014. “Updated Estimates of the Effects of the Insurance Coverage Provisions of the Affordable Care Act, April 2014.” [http://www.cbo.gov/sites/default/files/cbofiles/attachments/45231-ACA\\_Estimates.pdf](http://www.cbo.gov/sites/default/files/cbofiles/attachments/45231-ACA_Estimates.pdf).
- Duggan, Mark, Amanda Starc, and Boris Vabson.** 2014. “Who Benefits when the Government Pays More? Pass-Through in the Medicare Advantage Program.” National Bureau of Economic Research Working Paper 19989.
- Einav, Liran, and Amy Finkelstein.** 2011. “Selection in Insurance Markets: Theory and Empirics in Pictures.” *Journal of Economic Perspectives* 25 (1): 115–38.
- Ellis, Randall P.** 2008. “Risk Adjustment in Health Care Markets: Concepts and Applications.” In *Financing Health Care: New Ideas for a Changing Society*, edited by Mingshan Lu and Egon Jonsen, 177–222. Berlin: Wiley-VCH.
- Frogner, Bianca K., Gerard F. Anderson, Robb A. Cohen, and Chad Abrams.** 2011. “Incorporating New Research into Medicare Risk Adjustment.” *Medical Care* 49 (3): 295–300.
- Glazer, Jacob, and Thomas G. McGuire.** 2000. “Optimal Risk Adjustment in Markets with Adverse Selection: An Application to Managed Care.” *American Economic Review* 90 (4): 1055–71.
- Hall, Anne E.** 2011. “Measuring the Return on Government Spending on the Medicare Managed Care Program.” *The B.E. Journal of Economic Analysis & Policy* 11 (2).

<sup>32</sup> Currently, risk-adjustment formulae are calibrated using Marketscan data, but whether this sample mirrors the costs of the exchange population is unknown.

- Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7 (Special Issue): 24–52.
- Kim, You Suk, and Naoki Aizawa. 2013. "Advertising Competition and Risk Selection in Health Insurance Markets: Evidence from Medicare Advantage." Unpublished.
- Langwell, K., and J. P. Hadley. 1989. "Evaluation of the Medicare Competition Demonstrations." *Health Care Financing Review* 11 (2): 65–80.
- Lazear, Edward P. 2006. "Speeding, Terrorism, and Teaching to the Test." *Quarterly Journal of Economics* 121 (3): 1029–61.
- McWilliams, J. Michael, John Hsu, and Joseph P. Newhouse. 2012. "New Risk-Adjustment System Was Associated With Reduced Favorable Selection In Medicare Advantage." *Health Affairs* 31 (12): 2630–40.
- Mehrotra, Ateev, Sonya Grier, and R. Adams Dudley. 2006. "The Relationship between Health Plan Advertising and Market Incentives: Evidence of Risk-Selective Behavior." *Health Affairs* 25 (3): 759–65.
- Mello, Michelle M., Sally C. Stearns, Edward C. Norton, and Thomas C. Ricketts. 2003. "Understanding Biased Selection in Medicare HMOs." *Health Services Research* 38 (3): 961–92.
- Newhouse, Joseph P., Melinda Beeuwkes Buntin, and John D. Chapman. 1997. "Risk Adjustment and Medicare: Taking a Closer Look." *Health Affairs* 16 (5): 26–43.
- Physician Payment Review Commission. 1997. *The Physician Payment Review Commission's Views and Recommendations on Medicare HMO Payments*. Testimony of Gail R. Wilensky, Committee on Ways and Means, Subcommittee on Health, US House of Representatives, February 25, 1997.
- Pope, Gregory C., Randall P. Ellis, Arlene S. Ash, John Z. Ayanian, David W. Bates, Helen Burstin, Lisa I. Iezzoni, Edward Marcantonio, and Bei Wu. 2000. *Diagnostic Cost Group Hierarchical Condition Category Models for Medicare Risk Adjustment: Final Report*. Prepared for Health Care Financing Administration, December.
- Pope, Gregory C., John Kautter, Randall P. Ellis, Arlene S. Ash, John Z. Ayanian, Lisa I. Iezzoni, Melvin J. Ingber, Jesse M. Levy, and John Robst. 2004. "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model." *Health Care Financing Review* 25 (4): 119–41.
- Town, Robert, and Su Liu. 2003. "The Welfare Impact of Medicare HMOs." *RAND Journal of Economics* 34 (4): 719–36.
- Van de ven, Wynand P. M. M., and Randall P. Ellis. 2000. "Risk Adjustment in Competitive Health Plan Markets." In *Handbook of Health Economics*, Vol. 1., edited by Anthony J. Culyer and Joseph P. Newhouse, 755–845. Amsterdam: Elsevier.
- Yabroff, K. Robin, Elizabeth B. Lamont, Angela Mariotto, Joan L. Warren, Marie Topor, Angela Meekins, and Martin L. Brown. 2008. "Cost of Care for Elderly Cancer Patients in the United States." *Journal of the National Cancer Institute* 100 (9): 630–41.