Will risk-adjustment decrease health care costs? New evidence from the Medicare Advantage Program*

Jason Brown
U.S. Treasury Department

Mark Duggan
Univ. of Pennsylvania and NBER

Ilyana Kuziemko
Princeton and NBER

William Woolston
Stanford University

August 21, 2011

Abstract

To combat adverse selection, governments increasingly base payments to health plans and providers on enrollees' scores from risk-adjustment formulae. 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 "over-priced" individuals with costs significantly below the formula's prediction. We show that after Medicare risk adjusted capitation payments to private Medicare Advantage plans, plans enrolled individuals with higher scores but significantly lower costs conditional on their score, and overpayments to plans actually increased. Our results have implications for many cost-control reforms that use risk adjustment.

*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, Elena Nikolova, Christina Romer, Shanna Rose, Karl Scholz, Jonathan Skinner, and Luke Stein and seminar participants at Brown, Columbia, Cornell, Harvard, Houston, Northwestern, MIT, Princeton, RAND, Rice, Stanford, USC, U.S. Treasury Department, Wharton, Wisconsin, Yale, and the NBER Health Care meetings for helpful comments and feedback. Contact information for the authors: Jason.Brown@do.treas.gov, mduggan@wharton.upenn.edu, kuziemko@princeton.edu, william.woolston@stanford.edu. 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

1 Introduction

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 and increases health care costs—by paying providers or insurers fixed capitation payments rather than reimbursing them on the margin for each service. The success of such reforms hinges on correctly aligning capitation payments with a patient's expected cost. Overpaying an insurer or provider increases total health care costs, while differences in relative over- or underpayments can cause them to cream-skim the most over-priced cases instead of competing on quality or cost. To more accurately align payments to 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.

The Affordable Care Act of 2010 (ACA) relies heavily on risk adjustment in its attempt to both expand coverage and reduce the growth of health care costs. Approximately half of the more than 30 million people who are projected to gain coverage through the reform will do so via "insurance exchanges," in which private insurers will receive capitation payments partly based on an enrollee's health status. Outside of the exchanges, the law mandates that the entire individual and small-group markets be risk-adjusted by 2014. The ACA also increases the use of bundled payments and accountable care organizations by the Medicare and Medicaid programs; under a bundled payment system, providers who elect into these payment models receive a capitated, risk-adjusted payment for a certain condition such as diabetes or a hip fracture, rather than a payment for each procedure. Many European countries have undertaken similar reforms in an effort to decrease costs (Saltman and Figueras, 1998).

Despite the heavy reliance of risk adjustment in recent health care reforms, there has been limited empirical work assessing its effect on risk-selection or on the total cost of insuring a given population.[1] We provide the first assessment of the largest risk adjustment effort to date in the U.S. health care sector—Medicare's risk adjustment of capitation payments to private Medicare Advantage (MA) plans, which the ACA and subsequent regulation suggest as the model for risk-adjustment in both the state-run insurance exchanges and accountable

[1] There is a large, mostly theoretical or statistical, literature on risk adjustment, and 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.

care organizations—on the government's total cost of financing Medicare benefits.[2] 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.

Before 2004, capitation payments were, essentially, based on the cost of an average FFS enrollee with the same demographic characteristics and were not adjusted for health conditions. Even though open-enrollment regulations require 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 systematically more likely to enroll in an MA plan (Langwell and Hadley, 1989; Physician Payment Review Commission, 1997; Mello *et al.*, 2003; Batata, 2004). In response to this evidence of "differential payments" to MA plans—payments in excess of the actual cost of providing a given individual her Medicare benefits—in 2004 Medicare began to base capitation payments on an individual's "risk score," generated by a risk adjustment formula that accounted for over seventy disease conditions. We show that after 2004, differential payments to MA plans actually rise, and thus that risk adjustment increased the government's total cost of financing the care of Medicare enrollees.

How could risk adjustment, which was meant to decrease differential payments to MA plans, have made the problem worse? We offer a simple framework of risk-selection that can account for this unexpected finding and generates several predictions that we verify empirically. We first show that after risk adjustment, differential payments would have decreased had selection patterns into MA not changed. Before risk adjustment, MA plans had an incentive to enroll individuals who were low cost, both along dimensions that will later be included in the formula and those that will not. Because risk adjustment increases payments for individuals with the conditions included in the formula and decreases payments for those with few or no conditions, risk adjustment lowers the payments MA plans would receive for these individuals, as they were selected to have low risk scores.

But in response to the new incentives created by risk adjustment, selection patterns into MA change. After risk adjustment, MA plans have less incentive to avoid individuals with the conditions included in the formula but have a greater return to enroll individuals who have low costs *conditional on their risk score.* Indeed, relative to individuals who remain in

[2] See *Section 1343 — Risk Adjustment* of the ACA legislation, which suggests that "criteria and methods" similar to the HCC model be used in the exchanges. See `http://www.federalregister.gov/articles/2011/04/07/2011-7880/medicare-program-medicare-shared-savings-program-accountable-care-organizations` for information on risk adjustment for accountable care organizations.

FFS, we show that MA enrollees' risk scores increase after risk adjustment, but their costs conditional on their risk score fall so much that, if anything, MA enrollees have lower total costs after risk adjustment.

Finally, our framework shows that while risk adjustment indeed decreases plans' return to positive selection along the dimensions included in the formula, this effect can be more than offset by the increase in the return to positive selection conditional on the risk score. The key insight is that because the variance of medical costs increases with the expected mean, it is easier to find individuals with high risk scores who have, say, costs $2,000 below their capitation payment than it is to find individuals who do not have a single documented disease condition who are $2,000 cheaper than predicted. To take but one example, pre-risk adjustment, Hispanics were roughly $800 cheaper 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 $4,000 cheaper than their (risk-adjusted) capitation payment. By incentivizing MA plans to enroll individuals with a greater number of serious conditions, risk adjustment also incentivizes them to enroll individuals with highly variable costs. Due to this increase in variance, the return to differentially enrolling specific groups—whether through targeted advertising or designing benefits packages that differentially appeal to certain people based on demographics or disease history—can increase after risk adjustment, and with it the total cost of the Medicare program.

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).

Our results have important implications for attempts to control health care costs outside of the Medicare program, given that many cost-control proposals that move away from FFS rely on risk adjustment to set capitation payments. While we agree that the FFS model incentivizes over-provision, our results highlight challenges associated with implementing alternatives that rely on capitation. Moreover, risk-adjusting capitation payments outside of Medicare may prove even more difficult than within the program. Medicare directly insures the FFS population and thus has access to their claims and cost data to calibrate a risk adjustment model, whereas there is no such "public option" in the exchanges. State risk

adjusters might have to extrapolate from the Medicare or Medicaid population or else ask for cost and claims data from the private plans themselves—which would have an incentive to over-report costs if they knew they would be used to determine future capitation payments.[3]

It is important to emphasize that this study is not a full welfare analysis of the Medicare Advantage program or risk adjustment, as we focus on the cost to the government and not consumer or producer surplus. However, due to the sheer size of government health costs, analyzing how reforms affect public health care expenditure is of interest in its own right. Given their current level and growth rate, controlling Medicare costs—which by 2035 are expected to comprise over a quarter of the primary federal budget, or between five and six percent of GDP—is probably the most important factor in improving the long-term fiscal position of the U.S. government.[4] We estimate that differential payments to MA plans in 2006 totaled about $30 billion, or nearly eight percent of total Medicare spending that year, significantly greater than official government estimates which assume that risk adjustment works perfectly.[5] The MA program also appears to be at a crossroads: while there has been long-standing concern among some government agencies regarding the fiscal impact of the Medicare Advantage program and indeed the ACA cuts payments to some plans, the House of Representatives recently passed a budget measure that would end FFS Medicare and have private plans cover all Medicare beneficiaries.[6]

The remainder of the paper is organized as follows. Section 2 provides background information on the MA program and the risk-adjustment formula Medicare currently uses to adjust capitation payments to MA plans. Section 3 presents the intuition and results from the model. Section 4 describes the data. Sections 5 and 6 present the empirical results on selection and differential payments, respectively. Section 7 explores potential mechanisms by which MA plans might be able to differentially select certain enrollees. Section 8 discusses the challenges to improving risk adjustment and Section 9 offers concluding remarks.

[3] Of course, not all cost-control proposals rely on risk adjustment. See Chandra *et al.* (2011), Blumenthal and Glaser (2007), and Baicker and Chandra (2005) on, respectively, comparative-effectiveness research, health information technology, and malpractice reform.

[4] Projections are taken from the CBOs June 2011 long-term budget outlook: `http://www.cbo.gov/ftpdocs/122xx/doc12212/06-21-Long-Term_Budget_Outlook.pdf`.

[5] Unless otherwise stated, all dollars amounts reported in the paper are adjusted to 2007 dollars using the CPI-U.

[6] The proposed ACA cuts to MA payments echoes the recommendations the Medicare Payment Advisory Commission. (See, for example, `www.medpac.gov/documents/mar09_entirereport.pdf`, as well as reports from earlier years. Though they estimate large "overpayments" to MA plans, their methodology assumes that an individual with a risk score of one in FFS has the same underlying expected cost as an individual with a risk score of one in MA, an assumption we provide substantial evidence against in Section 5).) On the other side of the debate, the House bill, authored by Representative Paul Ryan, ends FFS Medicare for new enrollees starting in 2022 and replaces it with vouchers to purchase private health insurance, as in the current MA program (see `http://www.cbo.gov/ftpdocs/121xx/doc12128/04-05-Ryan_Letter.pdf` for the Congressional Budget Officer's summary and preliminary analysis of the proposal).

2 Background on Medicare Advantage capitation payments and risk adjustment

Beginning in the 1980s, Medicare enrollees have had the choice between the traditional fee-for-service system (FFS) and private MA plans. Plans must accept all applicants, charge all enrollees the same premium, and provide benefits generally comparable to traditional Medicare, but otherwise are free to coordinate patient care and thus can have varying benefits, cost-sharing arrangements, and provider networks. Instead of reimbursing providers for MA enrollees' services, the Medicare program pays MA plans a fixed capitation payment to cover these costs, and plans are thus the residual claimants if actual costs are above or below the level of 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.[7]

The capitation payment to an MA plan for covering an individual is thus based on the estimated Part A and B payments had FFS Medicare covered her directly. The Centers for Medicare and Medicaid Services (CMS), the government agency that administers Medicare, has generally focused on adjusting payments along two dimensions of expected cost: individual attributes and geography.[8] Individual attributes are used to generate an individual-level *risk score*. This risk score is then multiplied by a county-level *benchmark*, and capitation payments for individual i in year t in county c are given by: $Capitation\,payment_{itc} = Risk\,score_{it} \times Benchmark_{ct}$.[9] Below we describe how the methodology for calculating risk scores has changed, as well as how county-level benchmarks have evolved.

[7]MA plans that provide prescription drug coverage receive a separate capitation payment in return. All of our analysis on the fiscal impact of MA plans considers only the payments made to plans for covering Part A and B services. We discuss in the Conclusion how future work might examine the effect of Medicare Part D on the demand for MA plans.

[8]Since the 1980s, the names of both the MA program and the agency that administers Medicare have changed. For simplicity, we will always refer to private Medicare plans as "MA plans" and to the agency that governs Medicare as "CMS."

[9]County-level adjustment changed slightly in 2006, when MA plans were required to submit bids to CMS to cover the Part A and B costs of the average-cost individual in a given county. These were not, however, competitive bids in the traditional sense: instead of comparing MA plans' bids to each other and basing payments on the lowest bid made, CMS based payments on the difference between a plan's bid and the statutory county benchmark and plans that bid below the benchmark would receive 75 percent of the difference to spend on additional benefits for enrollees. Plans that bid at or above the benchmark would receive the statutory benchmark. CMS estimates this reform lowered MA payments by 3.4 percent in 2006, and, as described in the data section, we adjust our empirical estimates accordingly (see `www.medpac.gov/documents/jun07_entirereport.pdf`.

2.1 Risk adjustment before 2004

Throughout the 1980s and 1990s, county benchmarks were generally set to 95 percent of county FFS costs, as it was believed that MA plans should be able to deliver services more efficiently than FFS. However, the 1997 Balanced Budget Act began to weaken the strict link between benchmarks and FFS costs, and beginning in the late 1990s benchmarks were raised beyond FFS costs in areas of low MA penetration in an effort to expand access to MA plans. By the end of 2003, benchmarks were roughly 103 percent of average FFS spending.[10]

During the 1980s and 1990s, CMS used a "demographic model" to generate individual-level risk scores, so-called because it included only demographic variables (gender, age, and disability and Medicaid status) as opposed to disease or health conditions. Then as now, CMS does not require MA plans to report cost or claims data (doing so might be seen as being in tension with giving MA plans' freedom to coordinate care as they deem most effective), so it used the FFS population to estimate how each of these demographic factors contributed to average costs. Essentially, CMS regressed total Part A and B payments made on behalf of an FFS enrollee on the above demographic variables (including many interactions). CMS found that the model explained only one percent of the variation in Medicare spending among the FFS population (Pope *et al.*, 2004). Given the difficulties in out-of-sample prediction, it is unlikely that the model explained any more of the variation among the MA population, which, of course, is the population whose payments are actually determined by the formula.

Because of the low predictive power of the demographic model, there was substantial scope for MA plans to target enrollees who were healthier—and thus cheaper—than the demographic model would predict. Indeed, research using data from this period have found consistent evidence of differential payments. Estimates 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.[11]

Federal policymakers reacted to this evidence by enhancing the risk-adjustment procedure. In 2000, CMS introduced the principal inpatient diagnostic cost group (PIP-DCG) model. Due to the lack of MA cost data, the model used each Medicare recipient's inpatient diagnoses (thus excluding outpatient and physician diagnoses) documented on FFS claims data to predict FFS costs the following year. As MA plans do not submit claims data, ap-

[10]See `www.medpac.gov/documents/mar09_entirereport.pdf` for a description of the general evolution of benchmarks since the 1980s and `www.medpac.gov/documents/Mar04_Entire_reportv3.pdf` for information on benchmarks in 2003.

[11]This range is taken from the estimates in Langwell and Hadley (1989), Physician Payment Review Commission (1997), Mello *et al.* (2003) and Batata (2004). Related research has found evidence of favorable selection into private Medigap plans during this period as well, which provide supplemental coverage to enrollees' in traditional Medicare (Fang *et al.*, 2008).

plying this model to the MA population required that MA plans submit "encounter data," which document an enrollee's diagnoses. CMS found that the PIP-DCG model could explain 6.2 percent of the variation in FFS costs.

Between 2000 and 2003, risk scores were calculated as a 90/10 blend of the demographic and PIP-DCG models: $Risk\,score = 0.9 * Demographic\,score + 0.1 * PIP\text{-}DCG\ score$. Thus, the introduction of the PIP-DCG model raised the portion of MA cost variation explained by risk scores from one to $0.9 * 1 + 0.1 * 6.2 = 1.5$ percent.

2.2 Risk adjustment after 2003

The difference between county benchmarks and FFS costs continued to grow after 2003. By 2009, benchmarks in the average county reached 118 percent of FFS costs.

In an effort to reduce differential payments to MA plans, in 2004 CMS introduced a more comprehensive risk-adjustment regime that is based on the hierarchical condition categories (HCC) model. Like the PIP-DCG model, the HCC model uses claims data from the FFS population to calibrate a model that predicts FFS costs in the following year, though the HCC model accounts for not just inpatient claims, but physician and outpatient claims as well. The model distills the roughly 15,000 possible ICD-9 codes providers list on claims into seventy disease categories, the most common of which are described in Appendix Table 1.[12] Initally, the HCC model was blended with the demographic model, with the HCC model accounting for 30, 50 and 75 percent of the total risk score in 2004, 2005, and 2006, respectively. From 2007 onward, risk scores have been based entirely on the HCC model.

CMS found that when FFS data are used to calculate HCC scores, the HCC score explains eleven percent of FFS costs the following year (Pope *et al.*, 2004). Newhouse *et al.* (1997) and de ven and Ellis (2000) survey the literature and conclude that the lower bound on the percent of cost variation plans are able to predict is between 20 and 25 percent, suggesting there is still potential room for risk selection even if the model were to perform as well on the MA population as it does on the FFS population. Moreover, both prospective reports commissioned by CMS in 2000 and 2004 (Pope *et al.*, 2000 and Pope *et al.*, 2004) and more recent work using data from 2004 to 2006 (Frogner *et al.*, 2011) have found that the formula systematically under-predicts costs for those with the most serious health conditions, a fact we return to in Section 5.

Evaluation and application of the PIP-DCG and HCC models is complicated by the lack

[12] CMS provides the file mapping ICD-9 conditions to HCC categories at `http://www.cms.gov/MedicareAdvtgSpecRateStats/Downloads/RAdiagnoses.zip`. The model coefficients and algorithms can be found at `http://www.cms.gov/MedicareAdvtgSpecRateStats/Downloads/HCCsoftware07.zip`. County benchmarks are published annually in the Medicare Advantage "ratebooks" and ratebooks from 1990 to 2011 are all available at: `http://www.cms.gov/MedicareAdvtgSpecRateStats/RSD/list.asp`.

of cost or claims data from MA plans. Whether the model performs as well on the FFS population as it does on the MA population depends on at least two key assumptions: first, that that the coding practices MA plans use in generating encounter data to record the conditions used in the formula are the same as those used by FFS providers on claims data; second, that differences in the MA and FFS populations can be fully accounted for by these conditions.

CMS has done extensive research related to the first assumption. They have found that MA plans exhibit greater "coding intensity" in documenting disease conditions, so that an MA enrollee's risk score grows substantially faster than an FFS enrollee's risk score.[13] As risk scores are based on disease conditions from the previous year, the risk score of an enrollee the year he switches from FFS to MA is based on FFS provider claims data and is thus free of intensive coding. However, in subsequent years, risk scores are based on MA plans' coding practices, which are far more aggressive than those assumed when the HCC model was calibrated.

The HCC model will also offer an upwardly biased estimate of the counterfactual FFS costs of MA enrollees if these enrollees are positively selected along dimensions not included in the model. In fact, the introduction of risk adjustment will incentivize firms to selectively target individuals who they expect to have low costs *conditional on their risk score*. The next section describes a model for predicting how risk-adjustment will change selection patterns in MA plans as well as the total cost to the government of providing Medicare benefits to eligible enrollees.

3 Theoretical framework

In this section, we demonstrate that risk adjustment can actually increase the total cost to the government of providing a given public service. As noted in the Introduction, we focus on the cost to the government and not on consumer or producer surplus, though we return briefly to producer and consumer surplus at the end of the section. The formal model is relegated to the Appendix, and here we focus on the basic framework and make simplifying assumptions to better illustrate the intuition.

[13] See `www.cms.gov/MedicareAdvtgSpecRateStats/Downloads/Advance2009.pdf`. Note that this analysis does not mean that FFS providers are immune to the incentive to "up-code" diagnoses in order to increase reimbursements, a practice documented by Silverman and Skinner (2004) and Dafny (2005), but merely that they do not do so as intensely as MA plans. The CMS study focuses on the difference in growth rates between those who stay in FFS for at least two years in a row and those who stay in MA for at least two years in a row, in order to eliminate the effect of compositional changes. One reason that MA risk scores might grow faster is that the health of MA "stayers" is deteriorating faster than that of their FFS counterparts. However, CMS explicitly dismisses this possibility in a later analysis (see `www.cms.gov/MedicareAdvtgSpecRateStats/Downloads/Advance2010.pdf`).

In a system such as Medicare, where the government directly provides insurance but also finances competing private plans, the total cost to the government is the sum of (1) the direct cost of providing the guaranteed set of benefits to individuals who choose to remain in fee for service and (2) capitation payments to private firms for individuals who instead choose to receive these benefits via a Medicare Advantage plan. As such, to determine whether a capitation payment policy increases or decreases the government's cost of providing the guaranteed set of benefits, it is sufficient to determine whether it increases or decreases *differential payments*—the payments the government gives private plans for providing MA enrollees their Medicare benefits minus the cost had the government directly covered them.[14]

We call an individual's "actual cost" the cost to the government had she been covered by FFS, and though it is not necessary, assume here that the firm's cost of providing the basic FFS benefits package is the same as the government's.[15] To satisfy the individual's participation constraint, firms must spend at least as much as the individual's actual FFS cost or else she would prefer to remain in FFS. The greater the difference between the capitation payment and the FFS cost, the greater the potential surplus the firm and the "over-priced" consumer can split.

3.1 Modeling risk-adjustment

For simplicity, assume here that without risk-adjustment the government pays firms a fixed, constant payment per individual equal to the average FFS cost, and that after risk adjustment this payment varies depending on individual-level characteristics included in the risk-adjustment formula.[16] We further assume that the government employs a risk-adjustment

[14]This framework assumes that the cost to the government of directly covering a beneficiary is independent of the composition of the population enrolled in MA. But higher MA penetration might incentivize cost-control practices among providers, and these practices could spill over to how they treat their FFS beneficiaries. In this case, increased MA penetration could decrease total Medicare expenditures, even in the presence of positive differential payments as defined above. Using data from 1994 - 2001, Chernew *et al.* (2008) finds support for this hypothesis. However, using more recent data from 1999 - 2004, Nicholas (2009) finds no evidence of these cost-reducing spillovers.

[15]Whether, holding selection constant, private plans or the government can more efficiently provide a given individual with the basic benefits package certainly affects the total cost of the Medicare program. However, it does not affect the propositions of the model: how selection reacts to a change in risk adjustment and the effect of risk adjustment on differential payments. Whether the HMO model is actually more efficient than the traditional government fee-for-service model even absent selection effects is an open question. Duggan (2004) finds that when some California counties mandated their Medicaid recipients to switch from the traditional FFS system to an HMO, costs increased by 17 percent relative to counties that retained FFS. As, within a county, individuals did not select between FFS or an HMO, selection issues are unlikely to be driving the result.

[16]The results hold if the government pays firms more or less than the average FFS payment. Also, in the Appendix, we show that all results hold for incremental increases in risk adjustment. We focus here on the special case of an increase from no risk adjustment to some risk adjustment for the sake of simplicity.

formula that is "reasonable" in two ways. First, the correlation between the risk score generated by the formula and actual costs had the individual been covered by FFS is positive, and thus greater than the correlation between the non-risk-adjusted payment and actual costs (which is zero). Second, we assume that risk-adjustment is "payment-neutral"—that is, if the entire Medicare population joined MA the average payment is the same with and without risk-adjustment.[17]

3.2 Predictions regarding selection

For convenience, we define someone as being positively selected along the "extensive margin" if he has few of the conditions included in the risk-adjustment formula and thus a low risk score. An individual is positively selected along the "intensive margin" if he has low costs along dimensions excluded from the risk score. Someone positively selected along the extensive margin will have a low risk score and thus on average low actual costs, whereas someone positively selected along the intensive margin will have low actual costs *conditional on his risk score*.

Without risk-adjustment, firms are paid a fixed amount per enrollee and will thus seek out the individuals with the lowest expected cost conditional on their demographic characteristics—whether along the extensive or intensive margin—subject to the cost of encouraging them to enroll. These enrollment costs can vary by individual and might include the costs of devising benefits packages that appeal to one group but not another, targeted advertising and recruiting, or even the risk of sanctions due to violating open-enrollment regulations.

Before risk-adjustment, individuals with the conditions included in the formula represented likely losses for firms, since they would typically have above-average costs. Risk-adjustment, in contrast, makes the payment a function of the risk score and thus increases payments for individuals with the conditions included in the formula. For example, before risk-adjustment, enrolling people with breast cancer represented an expected loss because they have above-average costs. After risk adjustment, however, individuals with breast cancer experienced an increase in their capitation payment because "breast, prostate, colorectal and other cancers and tumors" is one category included in the HCC formula, so firms have a greater incentive to encourage these patients to join. More generally, risk-adjustment reduces the incentive to positively select along the extensive margin by avoiding individuals with the conditions included in the formula. Thus, we predict that after risk-adjustment, *the risk scores of those enrolling in MA will increase relative to those remaining in FFS.*

[17] As described in Section 2, introduction of the HCC formula indeed increased the correlation between the capitation payment and the actual FFS costs. We further assume payment neutrality so that our result that risk-adjustment can increase the government's cost is not driven by the government merely handing firms larger capitation payments on average after risk-adjustment.

While firms' incentive to avoid individuals with the health conditions included in the formula decreases after risk adjustment, the incentive to positively select along the intensive margin by differentially enrolling individuals with low expected costs *conditional on their risk score* increases. Returning to the example in the previous paragraph, before risk adjustment, firms may have wished to avoid all breast cancer patients. After risk adjustment, the capitation payment of a breast cancer patient will be based on the average cost of all people in the "breast, prostate, colorectal and other cancers and tumors" category. Empirically, individuals with breast cancer appear to be substantially less expensive than individuals with prostate or colorectal cancer, so these individuals may now be quite attractive to MA firms.[18] We therefore predict that after risk-adjustment, *actual costs conditional on the risk score will fall among those enrolling in MA relative to those remaining in FFS.*

3.3 Predictions regarding differential payments

While the selection predictions provide important tests for the model, they do not directly speak to whether risk adjustment decreases differential payments and thus the government's cost. In the formal model in the Appendix we show that, whenever risk-adjustment is "reasonable" in the sense defined earlier, it will decrease differential payments so long as selection patterns into MA do not change. We thus predict that *applying the risk-adjustment formula to the pre-risk-adjustment population of MA enrollees would have decreased the total capitation payments the government would have made on their behalf.*

However, the earlier selection predictions indicate that selection into MA will indeed respond to the incentives embedded in the formula, and in the model we show that the endogenous response to the formula can completely "un-do" the fiscal advantages of risk adjustment and actually increase differential payments. The key insight is that the variance of medical costs tends to increase with the expected mean and thus by incentivizing firms to enroll patients with a greater number of medical conditions, risk-adjustment also incentivizes them to enroll a population with more *variable* costs conditional on their risk score. This distributional characteristic of medical costs is well known (see, e.g., Lumley *et al.*, 2002), and can be seen in Figure 1, which plots actual costs as a function of (binned) risk scores, using our Medicare data. The difference between the mean and the 10^{th} percentile for risk scores between zero and 0.25 is \$2,000; the corresponding difference for risk scores between 2.75 and 3.25 is an order of magnitude larger.

That the variance of medical costs increases with their expected mean suggests that it should be easier to find, say, a breast cancer patient whose costs are \$1,000 below the average costs for people with breast, prostate, or colorectal cancer than finding an individual without

[18]See Yabroff *et al.* (2008). We return to the breast cancer example in Section 7.

a single documented disease condition whose costs are $1,000 below expectation. By incentivizing firms to positively select less along the extensive margin and thus enroll individuals with higher risk scores, risk-adjustment can increase the ability of firms to positively select along the intensive margin by attracting individuals with low costs relative to their capitation payments. Thus, *the effect of risk adjustment on differential payments is ambiguous.* It becomes an empirical question, which we explore in Section 6.

3.4 Discussion

Even in the more formal treatment in the Appendix, we do not specify how firms are able to attract "over-priced" consumers. Strictly speaking, how they do so is largely irrelevant to the government's bottom line. Firms might engage in targeted marketing, more slowly reply to the enrollment requests of individuals they suspect are less profitable, or design benefits packages in such a way that especially appeals to more profitable customers. Regardless of the actual selection mechanism, however, the government is overpaying the firm relative to the actual cost of supplying the baseline benefits package. As a more practical matter, knowing the selection mechanism is useful if the government wishes to reduce firms' ability to select over-priced enrollees, and we explore potential mechanisms in Section 7.

Finally, we are also rather silent on firm competition, as we believe it is likely not a first-order factor in determining the cost to the government. First, MA capitation payments are set by the government's risk-adjustment formula and not by competitive bidding. Thus, competition has no effect on the pricing schedule the government sets. Second, the most obvious effect of competition—how the differential payment is divided between firm profits and consumer surplus—does not affect the government's bottom line. In a highly competitive MA market, one would assume that to attract customers with a large difference between their capitation payment and actual costs, firms would likely pass on much of the difference to the consumer in the form of extra benefits.

But only the size of the differential payment—not how it is split between consumers and insurance plans—determines the government's cost. How market structure affects the division of the surplus between firms and consumers does not have a direct bearing on the government's cost, so we leave the question to future research.[19]

[19] Past studies have explored consumer surplus and profits in the MA market, as well as competition, 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 firm profits. Lustig (2010) estimates a structural model, which suggests that in 2002-2003, the welfare loss from imperfect competition in the MA market was greater than that from adverse selection.

4 Data

Our empirical work relies on individual-level data from the Medicare Current Beneficiary Survey (MCBS) Cost and Use series from 1994 to 2006. The MCBS 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 half a million claim-level observations each year.

Each year the MCBS follows a subsample of respondents for up to four years, thus creating a mix of cross-sectional and panel data. During our sample period, the data comprise over 55,000 unique individuals and over 150,000 person-year observations. From this sample, we make only minimal restrictions. First, we do not include the less than 0.25 percent of enrollees whose Medicare eligibility is based entirely on having end-stage-renal disease, as different risk adjustment and MA-eligibility rules apply to them. Furthermore, we exclude the roughly two percent of person-year observations who join the survey in the middle of the year.[20]

Appendix Table 2 reports summary statistics, separately for MA and FFS enrollees. For the purposes of this table, we follow CMS and classify an individual as an MA enrollee in a given year if he spends at least half of his Medicare-eligible months in an MA plan, though in most of the empirical work we make use of the fact that we know the number of months an individual is enrolled in MA each year.[21] FFS enrollees are more likely to be on Medicaid or disabled, and, conditional on not being disabled, are roughly a year older. While there are no significant differences with respect to gender, MA plans attract a disproportionate share of Hispanics and, to a lesser extent, blacks, likely reflecting the fact that MA enrollment is concentrated in urban and suburban areas and more limited in rural and other non-metro areas. The share of MA respondents with annual income above $20,000 (roughly the median of the sample) is about three percentage points higher than that of FFS respondents.[22]

4.1 Calculating risk scores and capitation payments

Some of the main predictions from the theoretical framework in Section 3 involve enrollees' risk scores, which are not included in the MCBS. We obtained risk scores from 2004 to 2006 for all MCBS respondents directly from CMS. However, testing the predictions also involves knowing what pre-period individuals' risk scores *would have been* had the HCC formula been in place, which of course CMS had no reason to document and which we therefore must

[20] The MCBS refers to these individuals as "ghost enrollees" and imputes some of their data.

[21] While enrollees can switch mid-year, well over 90 percent of individuals we classify as being enrolled in MA in a given year spend all twelve months in MA.

[22] We do not report the average income because the survey topcodes above $50,000.

generate ourselves. We are greatly aided in this task by the fact that an individual's risk score in year t is based on diagnoses documented on claims from year $t-1$ and that the MCBS includes all claims data for individuals in FFS. As such, using CMS's algorithm for converting claims data into risk scores, we can simulate the risk score for all MA enrollees the year immediately after they switch from FFS, as in that year their risk scores are based on FFS claims data that we observe. Given that we know the actual risk scores from 2004 to 2006, we can check the success of our simulation in these years: the correlation between our simulated risk scores and CMS's risks scores is over 0.96.

Another key variable in exploring the predictions of the framework is *Total Medicare expenditure*, the total cost to Medicare for individual i in year t, whether it is covering her directly via FFS or paying an MA plan to cover her. We calculate this variable by summing the reported capitation payment each month an individual is in MA and any Part A or B payments incurred over the year. Obviously, for those classified as being in MA, *Total Medicare expenditure* is determined entirely or mostly by capitation payments, and for those in FFS it is determined entirely or mostly by provider payments. Note that as the differential payment results from the model refer to how risk adjustment changes the total cost to the government, the *Total Medicare expenditure* variable refers only to costs incurred by the Medicare program and excludes individuals' out-of-pocket costs, though we return to out-of-pocket costs briefly in Section 7.[23]

While summing capitation payments and Part A and B payments is in principle very simple, another limitation of the MCBS is that, perhaps for confidentiality reasons, capitation payments reported after 2003 do not consistently reflect individual HCC scores.[24] However,

[23]One possibility we do not consider is whether changes to an individual's total Medicare expenditure affect his use of other government programs (federal or otherwise), and thus our results cannot speak to the effect on global government budgeting. For example, for individuals "dually" eligible for both Medicare and Medicaid, a decrease of $1 in the generosity of Medicare benefits would likely save the government less than $1 because such individuals would then rely more on the cost-sharing benefits provided by Medicaid.

[24]Two pieces of information support this conclusion. First, after risk adjustment when capitation payments were based on an individual's risk score, there is extremely little variation in the capitation payments recorded in the MCBS for beneficiaries in the same age group, calendar year, gender, disability, Medicaid status, institutional status, plan, and county cells. For example, in 2004 and 2005, consider all individuals who are (1) enrolled in an MA plan in May of that year and (2) are in a cell (as defined above) with at least one other beneficiary in the MCBS. Of these more 1,000 individuals, more than 92 percent have capitation payments that are within $1 of all other individuals in their cell. Second, using the actual risk scores provided to us by CMS, we show that individuals in the same cell (as defined above) who have *different* risk scores are recorded as receiving the *same* capitation payment. In 2006, the MCBS does not include plan identifiers, but the payment variable in the MCBS still does not appear to represent the actual amount of money an MA plan received. For example, there are twelve individuals who are enrolled in MA all months in 2006 and have exactly the same very low annual capitation payment ($913.58). Yet these individuals have substantially different risk scores (one has a risk score of 1.03 while another has a risk score of 4.67) and different ages (one is 68 years old while another is 95). We speculate that the MCBS may not include capitation payments that reflect an individual's risk score because such information would allow researchers to back out an individual's

because we have individuals' risk scores and we know their county benchmarks, generating capitation payments after 2003 is straight-forward. The final two rows of Appendix Table 2 show how the *Total Medicare expenditure* variable varies by MA status. The first version of the variable uses the risk scores to calculate capitation payments, and by this measure those in FFS cost Medicare over $500 less than do those in MA. The final row uses the uncorrected version of the capitation payment in the MCBS, with very similar results, suggesting that while the MCBS's capitation payment variable does not vary with individuals' risk scores in later years, it is roughly correct on average.[25]

The need to calculate HCC scores in the pre-period means we generally limit our analysis to those individuals who were in FFS all twelve months of a baseline year, so that we can be sure we have their complete claims history that year.[26] As such, much of the empirical strategy focuses on transitions from FFS to MA. Appendix Table 3 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 (these individuals often serve as a control group), and how these numbers change across our sample period. We observe over 1,500 individuals who switch from FFS to MA, and over 70,000 who remain in FFS over the course of two years.

Given that we often rely on those switching from FFS to MA for identification, we use a long pre-period in order to examine a substantial number of transitions. As discussed in Section 2, while there were some changes in MA policy between 1994 and 2003, namely the 1997 Balanced Budget Act, the major reforms were to the setting of county benchmarks, not the calculation of individual-level risk scores. However, as a robustness check, we show the magnitudes of our results are generally unchanged when we limit the sample to years after 1997.

---

risk score, a variable that is not included in the MCBS and that we needed to access directly from CMS itself. Nonetheless, as we show in Table 2, using the uncorrected capitation payments from the MCBS has little impact on our results.

[25]In practice, using the uncorrected MCBS capitation payment has minimal effects on our results, but given the problems we document in footnote 24 we think researchers should exercise caution when using this variable.

[26]As claims data are not available once an individual is in MA, we cannot do this imputation in subsequent years. As we discuss later, given the evidence described in Section 2 on "intensive coding" by MA plans, limiting our analysis to the change in Medicare expenditure the first year an individual switches to MA likely understates any increase in differential payments. Note that limiting the sample to those in FFS all of the previous year means we do not examine individuals the small share of individuals who join MA directly upon becoming Medicare eligible. For such individuals, who have no claims history upon which to base an HCC score, the original demographic model is used to calculate their risk score the first year they are in MA, with MA "encounter data" used in subsequent years.

5 How did selection patterns into MA change after risk adjustment?

In this section we test the predictions from the framework in Section 3 regarding selection into MA plans after risk adjustment. The framework offers two testable predictions regarding selection. First, after risk adjustment, extensive-margin selection should fall, thus leading to an increase in risk scores, at least for values of risk scores that are profitable in expectation. Second, intensive-margin selection should increase; conditional on risk scores, total Medicare expenditure for those switching to MA relative to those staying in FFS should fall after risk adjustment.

5.1 Quantifying the selection incentives created by the HCC model

As discussed earlier in Section 4, we focus our analysis on those in FFS for all twelve months of the baseline year, which includes many who will switch to MA the following year. For each individual, we calculate two counterfactual capitation payments were they to indeed switch to MA the following year: the first based on the original demographic formula and the second based on the HCC formula.

The first column of Appendix Table 4 compares these two capitation payments. The table includes only those in the pre-risk-adjustment period, before any selection endogenous to the HCC risk scores would be incentivized. Col. (1) presents the average difference between the HCC-based capitation payment and the demographic-based capitation payment, grouped by percentiles of the HCC score. Mechanically, capitation payments must, on average, rise under the HCC formula for those with higher risk scores, and col. (1) merely presents the magnitudes. For example, the HCC capitation payment would, on average, pay \$2,993 less than the demographic-based capitation payment for individuals with HCC scores in the lowest quarter, but it would pay more than \$29,000 more for the individuals in the top one percent.

Col. (1) would make it seem as though plans would be incentivized to increase risk scores over the entire risk score distribution, but col. (2) shows that doing so would not always be profitable on average. For example, individuals with the highest one percent of risk scores represent, on average, a \$6,000 loss to an MA plan, consistent with the work cited in Section 2 showing that HCC capitation payments for FFS enrollees with the most serious disease conditions would not fully cover these individuals' FFS costs. Thus, increasing risk scores indiscriminately would lead plans to enroll many beneficiaries who would be highly unprofitable in expectation. While positive profits might still be possible if these individuals were very positively selected along the intensive margin, these results suggest that firms might be reluctant to draw from the extreme right tail of the risk-score distribution.

5.2 Empirical strategy

To test whether firms react to the extensive-margin incentives depicted in Appendix Table 4, we estimate the following specification on the sample of individuals who are in FFS all twelve months of a given year t:

$$Risk\,score_{it} = \beta MA_{i,t+1} \times After\,2002_t + \gamma MA_{i,t+1} + \delta_t + \epsilon_{it}, \tag{1}$$

where i indexes the individual, t the year, $Risk\,score_{it}$ is the individual's risk score using year t claims data to predict Medicare expenditure in year $t+1$, $MA_{i,t+1}$ is the share of her Medicare-eligible months that the individual spends in MA in year $t+1$, $After\,2002$ indicates the baseline year is after 2002, and δ_t is a vector of year fixed effects.[27] This specification tests the prediction that after risk adjustment, the risk scores of individuals switching to MA will rise relative to those remaining in FFS. Note that while risk adjustment begins in 2004, those in FFS in the baseline year of 2003 would be switching into MA the first year under the HCC model, and are thus part of the "post-period."[28]

According to CMS' analysis of the risk formula (Pope *et al.*, 2004) and Appendix Table 4, on average capitation payments may not cover actual costs in the extreme right tail of the risk-score distribution. As such, we expect that firms will attempt to increase risk scores more in the middle part of the distribution than in the right-tail of the distribution, a prediction we test using quantile regressions.

To test the intensive-margin prediction, we estimate the following equation:

$$Expenditure_{it} = \beta MA_{i,t+1} \times After\,2002_t + \gamma MA_{i,t+1} + \lambda Risk\,score_{it} + \delta_t + \epsilon_{it}, \tag{2}$$

[27]While we could potentially use the administrative risk scores provided by CMS for the post-period, throughout this section we use our simulated risk scores in both the pre- and post-periods so that any change in risk scores between the two periods cannot be driven by differences in how they are calculated. Using the 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. As our simulated risk scores likely have some measurement error, we attribute this difference to less attenuation bias when the administrative risk scores are used in the post-period. We thus believe that the specification that uses the simulated risk scores in the pre-period and the actual risk scores in the post-period biases the analysis toward finding the effects we predict. Therefore, to be conservative, we prefer using the simulated risk scores in both periods.

[28]As readers may recall from Section 2, Medicare phased in risk-adjustment: in 2004 30 percent of the capitation payment was based on the risk score, in 2005 50 percent, and in 2006 75 percent. Because the effects we estimate tend to grow over the post-period, if we replace $MA \times$ *Post-period* with $MA \times Share\ risk\text{-}adjusted$, the magnitude of the implied effect of risk adjustment on extensive-margin selection, intensive-margin selection and differential payments are all significantly higher than those reported in Tables 1 and 2. For the sake of being conservative and making the coefficients readily interpretable, we choose to simply compare outcomes before and after the introduction of risk adjustment in 2004. That these effects grow over time may reflect MA plans becoming better at differentially selecting relatively over-priced patients or may simply reflect that as the risk-adjusted share increases, so do the incentives for differential selection.

where $Expenditure_{it}$ is the total FFS expenditure for individual i in year t and all other notation follows that in equation (1). We predict a negative coefficient on the interaction term—conditional on the risk score, after risk adjustment those enrolling in MA should have lower Medicare expenditure relative to those remaining in FFS because they are more positively selected along dimensions excluded from the risk formula.[29]

5.2.1 Results

Figure 2 shows how the distributions of risk scores change after risk adjustment, for, respectively, those remaining in FFS and those switching to MA, where we define someone as switching into MA if they spend at least half of their Medicare-eligible months the following year on MA. While there is a slight increase in risk scores after risk adjustment among the FFS population, the increase among those switching to MA is substantially greater. As predicted, this upward movement is concentrated in the middle of the distribution: the median risk score among those switching to MA increases by nearly forty percent after risk adjustment. However, risk adjustment does not increase the far right tail of this distribution, and thus the increase in the mean is more muted. The lack of movement in the far right tail is not surprising, given that individuals in that part of the distribution remain unprofitable on average even after risk adjustment.

The first five columns of Table 1 report results from estimating variants of equation (1). The first column presents the results from a basic OLS regression. The coefficients suggest that while individuals switching into MA before risk adjustment had risk scores roughly 0.31 points lower than those remaining in FFS, risk scores of those switching into MA rise after risk adjustment is introduced, making up about a third of the difference.

As noted earlier, we expect the effect on the mean to be muted, because risk adjustment should not have increased the incentive to enroll individuals in the far right-tail of the risk-score distribution. Indeed, in col. (2), merely dropping observations with risk scores above the 99$^{\text{th}}$ percentile (calculated separately each year) increases the magnitude and statistical precision of the estimate. Estimating a median regression (col. 3) on the entire sample increases the coefficient by nearly a third, so that sixty percent of the pre-period gap in risk scores between FFS and MA is made up by the increase in MA risk scores after risk adjustment.

We take col. (3) as our preferred specification—as it accounts for the fact that we do not expect to see much movement in the extreme right tail of the risk-score distribution but also

[29] This specification is similar in spirit to the "unused observables" test of Finkelstein and Poterba (2006). Using the terminology of their framework, $Expenditure_{it}$ in equation (2) is an "unused observable" because it is positively related to future costs to the insurer but, conditional on a beneficiary's risk score, is not used to determine insurance premiums or capitation payments.

allows us to use the entire sample—and subject it to a number of robustness checks. First, col. (4) shows that the result holds even when we employ the noisier measure of MA status used in Figure 3: whether an individual is in MA for at least half of his Medicare-eligible months the following year. Indeed, conditional on being in MA for at least half of the year and in FFS all of the previous year, the average value for *Fraction of year on MA* is 0.83, so we expect the coefficients in this specification to be smaller.[30] Col. (5) shows that restricting the sample to data starting in 1997 does not change the results.

The next four columns of Table 1 investigate intensive selection and test the prediction that costs conditional on risk scores for those switching into MA relative to those staying in FFS should fall after risk adjustment. Col. (6) reports the results from estimating equation (2). After risk adjustment, individuals switching into MA are over $1,200 "cheaper" than their risk scores predict them to be.

Col. (7) shows that while the coefficient is predictably smaller in magnitude when the MA indicator variable is used instead, it remains negative and significant. Col. (8) shows that, as with the extensive-margin results, the coefficients of interest are unchanged when the more recent sample is used.

The fact that the main effect of MA in Cols. (6), (7), and (8) is close to zero suggests that differences in risk scores accounted for essentially all of the cost differences between those joining MA and those switching to FFS in the pre-period. As risk scores were designed to address exactly these differences and were obviously based on data from the pre-period, the result is not surprising. However, once risk adjustment is instituted, MA plans have an incentive to enroll individuals who have low costs conditional on their risk scores, and the coefficients on the interaction terms suggest that enrollment patterns follow exactly that pattern.

Finally, col. (9) examines selection patterns of MA enrollees with respect to overall cost by re-estimating equation (2) without including the risk-score control. Before risk adjustment, MA plans had a strong incentive to attract low-cost beneficiaries. The results from col. (9) suggest that those individuals joining MA before risk adjustment have costs roughly $2,850 below those who remain in FFS. The results summarized in Section 3 and Cols. (1) - (8)

[30]Note that the corresponding figure among all individuals in MA for at least half of given year (not just those who had been in FFS the previous year) is 0.97, which is not surprising as those just switching in may have done so after January. Throughout the sample period, the rules on enrollment and disenrollment in MA plans went through mild changes. Besides being able to change plans each year during the annual "open enrollment period" the previous November, before 2002, individuals could switch between MA plans and FFS at any point in the year, in 2002 only during the first six months, in 2003-2005 only during the first three months, and in 2006 during the first six months. Throughout the sample period, institutionalized MA enrollees could return to FFS or switch to a different MA plan regardless of the time of year. At least in the MCBS, these reforms do not correspond to any change in the number of months spent on MA conditional on spending any months at all in a given year (results available upon request).

yield ambiguous predictions with respect to the change in selection with respect to overall costs after risk adjustment. On the one hand, individuals with higher risk scores tend to be more expensive; however, after risk adjustment, individuals who join MA are substantially less expensive than their risk score suggest. Col. (9) shows that, if anything, individuals who join MA after risk adjustment are slightly more selected with respect to overall cost, though the result is not statistically significant.[31]

5.3 Discussion

The results in this section provide strong support for the predictions of the theoretical framework in Section 3 regarding selection patterns. First, firms have less incentive to engage in extensive-margin selection after risk adjustment, and as a result risk scores rise for those switching into MA relative to those remaining in FFS. Moreover, this increase only occurs in regions of the risk-score distribution that support positive expected profits, providing further evidence that firms are indeed reacting to incentives. Second, conditional on the risk score, those switching to MA have lower baseline FFS spending relative to those staying in FFS after risk adjustment, consistent with their being more intensely selected along dimensions excluded from the risk formula. In the MCBS data, these two effects appear to roughly cancel each other out, leaving the positive selection with respect to overall baseline costs essentially unchanged after risk adjustment.

Recall that our framework shows that if firms increase their intensive-margin selection efforts, differential payments can actually rise after risk adjustment. While risk adjustment will always decrease differential payments if selection patterns do not change, the impact on differential payments when selection patterns do change is an empirical question, which the next section explores.

6 Did risk adjustment decrease differential payments to MA plans?

In this section, we focus on how an individual's *Total Medicare expenditure* in a given year changes as he switches from FFS to MA. Recall from Section 4 that *Total Medicare expenditure* is equal to the total cost to Medicare for an individual, whether it is covering her directly

[31] One limitation of the results in Table 1 is that our data only cover the first three years of risk adjustment. We therefore look to county-level data through 2008 (the most recent year available) to see whether selection patterns during the 2004-2006 period look similar to those during the 2007-2008 period. The county-level data only include average per capita costs and thus do not allow us to test the specific predictions regarding extensive- and intensive-margin selection. Instead, we merely seek to show that with respect to selection on *overall costs*, MA enrollees appear similar before and after 2006. In the interests of space we do not present these results, but they can be seen in Table 4 of a longer, working-paper version of this study (Brown *et al.*, 2010). We find no evidence that selection with respect to overall costs change after 2006, and the p-value of the change in our preferred specification is 0.94.

via FFS or paying an MA plan to cover her. If risk adjustment works perfectly—so that in expectation capitation payments are equal to an individual's FFS costs—then whether an enrollee switches between FFS and MA should have no effect on his total Medicare expenditure levels.

To isolate the effect of the introduction of risk adjustment from other changes occurring around the same time, we make two adjustments to capitation payments after 2003. First, as noted earlier, the growth rate of county benchmarks (the baseline value, which, multiplied by the risk score, yields capitation payments) began to rise in the later years of our sample period, in some counties considerably so. We therefore calculate capitation payments holding the growth rate of each county's benchmark to its pre-2004 level. Second, in the years immediately following the introduction of risk adjustment, plans received a so-called "budget-neutrality" adjustment (about a ten percent increase in capitation payments) to ease the transition to risk adjustment, and we reduce payments to remove this effect. In both cases, these adjustments increased all capitation payments by a given percent and did not depend on underlying individual conditions or characteristics. The adjustments we make obviously decrease the likelihood we would observe an increase in differential payments after risk adjustment.[32] Before examining whether differential payments fell after the introduction of risk adjustment in 2004, we explore whether they would have decreased had selection into MA plans held to its pre-2004 patterns, a key prediction from Section 3.

6.1 Would risk adjustment have worked had selection patterns not changed?

Figure 3 overlays the distribution of the year-on-year change in *Total Medicare expenditure* for those switching from FFS to MA between 1994 and 2003 and the distribution of that change from the same population *had the HCC model been in effect*. On average during the pre-period, actual Medicare costs increase by \$3,049 for the MA joiners but by only \$1,163 for the FFS stayers. The unconditional differential increase in Medicare expenditure is therefore \$1,886. When we instead simulate the capitation payments under the HCC model for those switching to MA, differential payments shrink by roughly \$1,000. This \$1,000 decrease remains when we condition for the large set of control variables listed in the notes to Table 2.

[32]There is one reform for which we do not adjust, as doing so would make it more likely to find that differential payments increased after risk-adjustment. In 2006 MA plans began to submit "bids" to CMS, with plans bidding below their county benchmark receiving 25 percent of that difference to use to finance extra services for their enrollees. The MCBS does not record plans' bids and thus we cannot perfectly calculate an individual's capitation payment even knowing the risk score and benchmark in 2006. MedPAC estimates that this bidding reform reduced capitation payments to MA plans by 3.6 percent in 2006. To be conservative, we reduced capitation payments by five percent in 2006, as it is possible that the plans in the MCBS submitted below-average bids.

In short, had selection patterns not changed, we predict that the introduction of the HCC formula would have substantially reduced differential payments. Of course, given that the formula was calibrated on this population, this is a relatively undemanding test of the risk-adjustment model. Moreover, as Section 5 demonstrates, selection patterns changed substantially after risk adjustment—first, MA enrollees' risk scores increased once capitation payments became a function of risk scores, and, second, conditional on risk scores, pre-period FFS costs for this group fell. As we argued in Section 3, whether the change in selection patterns can completely "un-do" the risk-adjustment model is an empirical question to which we now turn.

6.2 Did risk adjustment reduce differential payments after 2003?

6.2.1 Empirical strategy

We begin, as usual, with the sample of beneficiaries in FFS all twelve months of a given year $t-1$. To estimate the counterfactual Medicare expenditure for an MA joiner in year t had he remained in FFS, we examine the actual Medicare costs in year t for FFS stayers who are similar along observable dimensions. The estimating equation is:

$$Expenditure_{it} = \beta MA_{it} \times After\,2003_t + \gamma MA_{it} + \lambda X_{it} + \delta_t + f(Expenditure_{i,t-1}) + \epsilon_{it}, \quad (3)$$

where $Expenditure_{it}$ is total Medicare expenditure for person i in year t, $f(Expenditure_{i,t-1})$ is a flexible function of lagged Medicare expenditure, and all other notation follows that in previous equations. We prefer this specification to simply regressing $\Delta Expenditure_{it}$ as the lagged expenditure controls in equation (3) can better account for the fact that medical costs typically exhibit strong regression to the mean, though results using $\Delta Expenditure_{it}$ look very similar and are available upon request.[33]

The coefficients β and γ estimate the change in total Medicare expenditure associated with an individual's switching to MA relative to his having stayed in FFS. These estimated effects are consistent only if MA is uncorrelated with ϵ. This condition implies that, conditional on our control variables, the decision to join MA is not systematically related to time-varying shocks to an individual's expected cost to the Medicare program. If, for exam-

[33]The lagged Medicare expenditure controls include: lagged Part A and B expenditure and deciles of non-zero Part A payments and non-zero Part B payments as well as indicator variables for zero Part A and B payments (we found that regression to the mean differed depending on the type and level of costs). Note that regressing the *change* in spending is thus nested in the equation (3)—the two are equivalent if the coefficient on lagged expenditure is constrained to equal one and the coefficients on all other lagged expenditure variables are constrained to equal zero. The results are not sensitive to controlling more coarsely or finely than deciles for lagged Part A and B expenditure.

ple, individuals join MA when their health is deteriorating and thus their expected costs are rising, the effect of MA enrollment on total Medicare expenditure will be positively biased. However, when we return to endogeneity concerns later in this section, we argue that they generally bias results against finding positive differential payments to MA plans.

6.2.2 Results

The first column of Table 2 shows the results from merely regressing the level of total Medicare spending on the MA variable, which is allowed to have a different effect before and after risk adjustment, the lagged spending variables, and year fixed effects. Total Medicare expenditure increases by roughly \$905 when an individual switches from FFS to MA (for the entire year) before risk adjustment, and by an additional \$1,733 after risk adjustment.

The second column adds county fixed effects as well as demographic and other basic controls (all listed in the table notes). The coefficient on the interaction term increases to \$2,268. These controls are important if, for example, older people tend to have higher spending growth and post risk adjustment they are also more likely to join MA plans. In this case, we want to account for the fact that these older beneficiaries would have likely experienced high cost growth had they remained in FFS. That the coefficient on the interaction term increases by nearly a third suggests, as we hypothesized earlier, that selection endogeneity works against finding MA differential payments, at least in the post-period.

Col. (3) includes measures of lagged health indicators, which has no effect on the coefficient on the interaction term. Col. (4) includes health indicators from the current year. We prefer this specification over col. (3) as it better accounts for potential regression to the mean in health status—if enrollees typically experience a deterioration in their health upon joining MA, then comparing current to previous year's spending will over-state MA differential payments. We are aware, however, that health status will be endogenous to the care individuals receive in MA versus FFS and thus including it may be "over-controlling." In practice, the two estimates are very similar.

Cols. (5), (6), and (7) subject the estimation in col. (4) to robustness checks. The coefficients predictably fall in col. (5) when the fraction of the year spent in MA is replaced by an indicator variable for having spent at least half of the year in MA, though both remain positive and highly significant. Windsorizing the data based on the 99$^{\text{th}}$ percentile in col. (6) or dropping years before 1998 in col. (7) leave the results largely unchanged. Appendix Table 5 replicates all of Table 2, but uses our simulated HCC scores to calculate post-period capitation payments instead of the administrative HCC scores. The point estimates are almost identical.

6.3 Calculating the total value of differential payments

To fully measure the fiscal impact of MA enrollment, we use the actual payments to plans in col. (8) of Table 2, including the budget-neutrality payments and allowing county benchmarks to grow at their actual rate. While the coefficient on the main effect remains essentially unchanged from that in col. (4), the coefficient on the interaction term grows substantially.

We use these results to estimate total differential payments in 2006, when MA enrollment was 7.6 million: $(\$767.5+\$3,257.6)*7.6$ million $=$ \$30.6 billion, or 7.9 percent of *total* Medicare spending on benefits in 2006.[34] Of course, given the standard errors in the differential payment regressions, our need to calculate capitation payments from the administrative risk scores instead of having administrative capitation payments, and the fact that the coefficients are identified by individuals switching from FFS to MA and not the entire Medicare population, this calculation should be taken as merely our best estimate. For the sake of comparison, in col. (9) we replicate this final estimate but "naively" use the uncorrected MCBS capitation payments, and obtain very similar results. Note that whatever the size of total differential payments in 2006, they have likely grown since that time, as the MA share of the Medicare population and underlying county benchmarks have both continued to grow over the past five years.

6.4 Discussion

Given that our identification relies on those switching to MA from FFS, a natural question is whether the cost differences for this group are representative of the differences between the stock of MA and FFS beneficiaries. For example, while readers may agree that we have indeed identified large differential payments the first year an individual is in MA, perhaps the cost-containment measures of MA plans slow cost growth thereafter, so that differential payments shrink after the initial year.

As documented earlier, CMS itself has provided evidence against this hypothesis.[35] Specifically, they have found that the growth rate of risk scores in MA is faster than that in FFS. They attribute this phenomenon to "intensive coding"—enrollees in MA plans being diagnosed more aggressively than their FFS counterparts. Because risk scores in the first year that a beneficiary is enrolled in MA depend on conditions recorded when he is still enrolled in FFS, the effect of intensive coding can begin no sooner than an individual's second year in

[34]Medicare spending on benefits (i.e., excluding administrative costs) totaled \$385 billion in 2006 (in 2007 dollars, to match the coefficient units). See `http://www.kff.org/medicare/upload/7305-02.pdf`, which gives nominal estimates that we adjust with the CPI-U.

[35]See the 2010 CMS Advance Notice to MA Plans, `www.cms.gov/MedicareAdvtgSpecRateStats/Downloads/Advance2010.pdf`.

MA. Given that the growth in risk scores is mechanically related to the growth in Medicare expenditure for the MA population, and that risk scores in the FFS population are, in expectation, equal to Medicare expenditure by the very construction of the risk score, this finding by CMS suggests that our estimates for new MA enrollees likely understates the average differential payments for the stock of all MA enrollees. After years of intensive coding on the part of MA plans, the difference between capitation payments and counterfactual costs in FFS should fan out further, not contract.

Turning to another potential concern, while we mentioned earlier that endogeneity in equation (3) likely works against finding positive differential payments to MA plans, readers may wish for further evidence. Any bias story working in the opposite direction must argue that while those switching to MA appear relatively healthy and low-cost the year before they switch, they systematically have higher medical costs their first year in MA and thus would have also been expensive had they remained in FFS. These individuals might experience deteriorations in their health just after they switch or perhaps they put off an expensive medical procedure until joining MA.

We find such stories unlikely for several reasons. First, if those switching from FFS systematically experience a deterioration in health their first year in MA, then controlling for current-year health as we do in col. (4) should have substantially reduced the coefficients on the MA variables. Second, individuals are unlikely to postpone expensive procedures until they join an MA plan because plans tend to have less generous cost-sharing arrangements for serious medical procedures than does FFS. In fact, we find that individuals who switch to MA were no less likely to have an eye exam their last year in FFS, even though vision coverage is generally more generous in MA and thus one might imagine those about to switch out of FFS would put off the exam until they enroll in an MA plan.[36] Finally, consistent with there being no "Ashenfelter dip" the year before a switch to MA, when we control for the past two years of medical costs, instead of just one as in Table 2, the coefficients on the MA variables barely change, though become significant at only the ten-percent level due to the sample size falling by a half.

Considering the incentives MA plans face, these facts are not surprising. The least profitable enrollees for them in the post-period would be those who have little contact with the medical system their last year in FFS—and thus no documented HCC conditions and thus a low capitation payment—but suddenly become expensive their first year in MA. As we have shown throughout the last two sections, plans seem able to enroll the most profitable

[36] On the general tendency of MA plans to have less generous cost-sharing for serious conditions, see the Kaiser Family Foundation's report on MA benefits, `http://www.kff.org/medicare/upload/8047.pdf`. Note that the MCBS asks about eye exams the past year but not other specific examples of medical check-ups.

Medicare beneficiaries, though both the empirical work and model has not specified exactly how they do so. The next section seeks to shed light on possible mechanisms.

7 How does selection into MA plans take place?

In this section we explore potential mechanisms underlying the selection results. We begin by exploring why low-cost individuals tend to be enrolled in MA plans in both the pre- and post-periods, and then focus on why enrollment patterns changed after risk-adjustment.

7.1 Why are low-cost individuals more likely to be in MA plans?

The evidence in Section 5 shows that MA plans enrolled substantially lower-cost individuals before and after risk adjustment. But how do such patterns emerge when, due to open-enrollment requirements, plans must offer the same plans at the same rate to all Medicare beneficiaries in their geographical area of operation? While one can imagine several different possibilities, here we explore whether once individuals enroll in MA, the healthy ones are differentially more satisfied. 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.[37]

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 calendar year or the previous calendar year as well. As such, for this section we generally sample those who did not switch (either from FFS to MA or from MA to FFS) the previous year. Thus, unlike the majority of analysis so far in the paper, identification comes from cross-sectional variation—comparing individuals in MA with individuals in FFS. Asking someone who, say, just switched from FFS to MA to rate their health care "last year" would likely shed little light on their experience so far in MA.

This sampling means we actually have little information on the medical spending of those in MA. Recall that after someone enters MA, the MCBS—and indeed the Medicare program itself—does not track their medical claims or costs, and without this information we cannot

[37]Though we do not have the data to investigate them, other mechanisms have been suggested by past work. Targeted advertising may play a role: A report by the Kaiser Family Foundation concludes that advertisements for MA plans target healthy people (see `http://www.kff.org/medicare/upload/7805.pdf`). Health plans may delay responding to individuals they suspect are high-cost: Bauhoff (2010) finds evidence via an audit study that highly regulated German health insurance firms respond more quickly to enrollment requests from respondents residing in low-cost areas of the country, and MA plans likely have far greater flexibility than do German firms.

compute risk scores. As such, we cannot test whether the specific extensive- and intensive-margin selection patterns also arise with respect to satisfaction. We would have liked, for example, to see whether MA plans treat individuals with low costs relative to their risk scores better after risk adjustment, but such detail is impossible given data limitations.

Instead, we focus on the fact that both before and after risk adjustment, MA plans enroll individuals who are far healthier than average. Before risk adjustment, MA and FFS enrollees in our regression sample have mean self-reported health (from one, "poor," to five, "excellent") of 3.36 and 3.11, respectively. The differential shrinks slightly, to 3.34 and 3.13, after risk adjustment, but the change is not close to being statistically significant. Given the results in Section 5 that selection along baseline total costs did not change, this result is not surprising.[38]

One way such a pattern could arise is that sicker enrollees are differentially unsatisfied in MA and are thus more likely to switch back to FFS. To test this hypothesis, we estimate the following equation:

$$Satisfaction_{it} = \beta MA_{it} \times Health_{it} + \gamma MA_{it} + \mathbf{H}_{it} + \lambda X_i + \delta_t + \epsilon_{it}, \qquad (4)$$

where $Satisfaction$ measures individuals' reported satisfaction with different aspects of their health care and varies from one (very dissatisfied) to four (very satisfied), $Health$ is the five-category self-reported health variable described earlier, $\mathbf{H}$ 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, and thus the interaction term captures 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 differently—for example, Hispanics rate their health lower than non-Hispanics, even though throughout our sample period they have lower health costs.

Table 3 displays the results from estimating equation (4) via OLS.[39] We demean the $Health$ variable in $MA \times Health$, so that the MA main effect represents the effect of 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—

[38]Of course, MA plans may have a treatment as well as a selection effect on enrollees' health. Aizer *et al.* (2007) shows that the switch to managed care in California's Medicaid program had deleterious effects on pre-natal care and infant health. Risk-adjustment could in fact decrease plans' incentives to maintain enrollee health, as they now receive additional compensation if enrollees develop any of the conditions in the formula.

[39]Note that ordered logit gives exactly the same patterns of coefficient signs and significance levels, and results are available upon request. The sample size variation across different regressions arises from variation in the number of individuals who report not having enough experience to make a satisfaction rating as well as some questions not being asked in the earlier years.

suggesting that someone of average health reports lower satisfaction in MA than in FFS. Of course, the type of person who joins MA might simply be harder to please—after all, FFS is the default and they chose to switch in the first place. As such, one should be cautious in interpreting this coefficient as proof that MA plans deliver poorer services on average.

We instead focus 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. 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. In five of the nine categories (overall, out-of-pocket costs, doctor's concern for your health, questions answered over the phone, and receiving information about your medical condition) the coefficient is significant, and a sixth (having medical care provided in the same location) has a p-value of 0.113.

The last row of Table 3 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, or to FFS if he is currently in MA—on the same set of explanatory variables. 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.[40]

7.2 Possible mechanisms underlying the changes after risk-adjustment

The results in Table 3 provide a potential explanation for why in general higher-cost enrollees wind up more often in FFS, but they do not explain how individuals with low costs relative to their risk score differentially found their way into MA plans after risk-adjustment. While we cannot provide a definite answer, we believe several factors are at work.

First, plans have a wealth of data, both on their own MA enrollees and from their operations in the non-Medicare market. In fact, because plans (unlike the Medicare program) have data on the medical claims and costs of their MA beneficiaries, along this dimension plans have more information than the risk-adjustor. They can use these data to, say, determine which diseases have the greatest variance conditional on their risk score and differentially

[40] One might assume that because those exiting MA do not appear particularly healthy relative to the FFS stock, our selection results are over-estimated and thus our differential payment results may be over-stated as well. However, differential payments are a function not only of selection but also of capitation payments, and CMS has documented that risk scores grow faster in the MA stock than in the FFS stock due to intensive coding. So, at the point when they return to FFS, MA enrollees' capitation payments would have grown faster than their actual costs, and thus differential payments are in fact under-estimated by considering only those switching from FFS to MA. See Section 6 for further detail.

attract individuals with those diseases. Moreover, given that demographic information such as income, education, and race is not included in the risk formula, firms might be able to determine that, say, Hispanics with a history of congestive heart failure are over $4,000 cheaper on average than their risk score suggests.[41] Given that the variance in costs grows with the risk score, demographic differences that are small in a relatively healthy population could be very large for a specific disease. Not only might *Demographic group* $\times$ *Disease category* have substantial explanatory power even conditional on the risk score, but the individual traits embodied in each term would seem to be easy to differentially target in advertising campaigns (e.g., advertise on a Spanish radio station that if you join Plan X you will never wait more than two days for an appointment with an cardiologist).

While we can use the entire FFS population in the pre-period to determine which groups are "over-priced," we have far less scope to test whether these same patterns are reflected in differential enrollment patterns into MA after risk-adjustment. Cutting the relatively limited number of individuals switching from FFS to MA by relatively specific criteria quickly runs into power constraints.[42]

Given these constraints, we can only examine how enrollment patterns change for relatively large demographic and common disease groups. Fortunately in terms of data analysis, one of the most common HCC categories is Category 10, "Breast, Prostate, and Colorectal Cancer," and given the prevalence rates of these diseases, the large majority of women in this category will have breast cancer while the large majority of men will have 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 (in our data, the difference is $1,200, though given the category also includes colon cancer we would expect some attenuation bias). Therefore, in this category, unlike the rest of the common categories, all of which do not aggregate distinct conditions (see Appendix Table 1), we can identify a group of individuals (women) who were underpriced before risk adjustment and are over priced 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

[41]We estimate this difference using the FFS population from before 2004, so that no selection with respect to the risk score should have yet taken place. Note that it is more difficult to estimate differences in overpayments with respect to the original demographic formula because individuals remaining in FFS are negatively selected with respect to the formula in both the pre- and post-periods.

[42]For example, among Hispanics in FFS, the share with a history of congestive heart failure is the same before and after risk adjustment, whereas that share goes from zero to thirteen percent among Hispanics in MA. And even though the corresponding trends for non-Hispanic MA enrollees are slightly negative, the relative increase in Hispanics MA enrollees with a history of congestive heart failure is not statistically significant as the absolute increase is only four individuals.

after risk-adjustment ($p < .02$). Given that the one *Demographic* $\times$ *Disease* category we can reasonably test behaves as expected, it gives some assurance that plans would indeed target their efforts at groups that offer even larger expected differential payments.

Second, beyond advertising campaigns that target groups newly over-priced under the HCC formula, plans may change their benefits, networks, and cost-sharing in such a way that low-cost individuals conditional on their risk score will select into them. Glazer and McGuire (2000) and Ellis and McGuire (2007) model how managed-care firms could design benefits packages that differentially appeal to healthy individuals. Extending their logic to explore the effect of a change in risk-adjustment, we hypothesize that firms wishing to attract individuals with higher risk scores but low costs conditional on the risk score might decide to offer greater access to specialists for individuals with serious diseases, while at the same time increasing cost-sharing. Such a plan would attract a cancer patient in remission who may need to see an oncologist a few times a year, but not someone in a more severe stage of the disease.[43] Indeed, after risk-adjustment, MA enrollees in self-reported poor or fair health report having improved access to specialists, relative to their FFS counterparts ($p < .05$). However, they experience no improved satisfaction with respect to cost-sharing. In fact, after risk-adjustment, MA enrollees regardless of self-reported health report greater dissatisfaction with cost sharing, relative to FFS enrollees ($p < .001$). If plans are offering targeted benefits to appeal to individuals with a history of certain serious diseases and thus a higher risk score, tighter cost-sharing may be essential to selectively attract individuals with low expected costs conditional on their risk score. We find these results at least suggestive that plans changed their cost-sharing arrangements in response to risk-adjustment.

While we feel the results in this section begin to shed light on how MA plans might actually risk-select, which our framework and even the more detailed model in the Appendix treated in a very reduced-form manner, given the data limitations mentioned earlier, we feel this topic warrants further work. Bauhoff (2010) uses an audit design in his work on German insurance firms, and applying his methods to MA plans could help researchers and policy-makers gain further understanding of the mechanics of risk-selection, as well as how it responds to risk-adjustment. Given that Affordable Care Act will mandate risk adjustment in the entire small group and individual insurance markets in 2014, research on how risk-selection actually takes place is likely to have even greater public-policy relevance in the near future.

[43] Among all the satisfaction questions in the MCBS, "access to specialists" seemed the most likely to differentially appeal to individuals with serious disease conditions. Healthy and sick people alike should appreciate, say, having questions answered over the phone.

8 Challenges associated with improving risk-adjustment

8.1 Recalibrating the model

Given evolution in medical treatments and research, the costs associated with diseases change over time. As such, the coefficients in any risk-adjustment model will need to be recalibrated.

However, the combination of intensive-margin selection and the lack of cost data once beneficiaries leave FFS makes recalibration especially difficult in the context of Medicare Advantage. Section 5 documented that after risk-adjustment, the individuals remaining in FFS have higher costs relative to their diagnoses than those switching to MA. Under the current data collection arrangements, when the government wants to re-estimate the costs for each disease, they can only do so on the FFS population. But because the sample will have been endogenously selected to contain the most expensive cases for each disease, the re-calibrated formula will produce upwardly biased estimates of the cost of any given disease.

Alternatively, recalibration might benefit from using cost data from individuals in the MA population, which, as noted earlier, CMS does not currently collect. If, for example, MA beneficiaries with certain conditions are systematically less costly than their risk score suggests, the government could reduce capitation payments for these conditions. However, basing a plan's capitation payments on costs incurred recreates the very marginal-cost reimbursement the MA program was designed to avoid. The "yardstick" model of Shleifer (1985) suggests that a plan's capitation payments could be based on the submitted cost data of *other* plans in its geographic area of operation, though he notes that collusion between plans could undermine the effectiveness of such a system.

8.2 Adding more categories to the formula

A natural reaction to the intensive-margin results in Table 1 is that the government should simply add more detail to the formula. Indeed, our results in the previous section comparing breast and prostate cancer suggests that the formula may be too parsimonious. However, there are several drawbacks to simply adding more categories.

First, doing so could provide firms more scope to "intensively code." If the only two conditions in a formula are heart attack and cancer, outside of actual fraud, MA plans cannot document that a patient has one of these conditions when he in fact does not. But "diabetes with complications" is far more open to interpretation; CMS reports that, relative to FFS, MA plans tend to interpret gray areas in a manner that results in higher risk scores.

A less obvious drawback is that having a more flexible HCC model would mean increasing the number of estimated coefficients for the risk-adjustment model. While the sample of individuals on which CMS can perform its estimation is large, it is not unlimited. When

choosing the complexity of a risk-adjustment model, CMS faces a tradeoff between adding more parameters to their model (and explaining more of the variance in costs) and measuring each coefficient precisely. Even if each coefficient is unbiased, having mismeasured coefficients can lead to large differential payments if MA plans are able to estimate the model more precisely than is CMS and attract patients with the over-priced conditions.[44]

Finally, certain highly predictive factors may be impossible to include in the model for either practical or political reasons. The most obvious example is prior year's costs—this variable is in fact more predictive of current year costs than is the HCC risk score. However, plans cannot be compensated based on prior year costs or else they would have an incentive to increase costs this year so as to receive a larger capitation payment the following year, meaning that firms would be paid at the margin for services performed, albeit with a one-year lag. Race and ethnicity are also highly predictive, but perhaps for reasons of political sensitivity, Medicare has never used these factors to adjust prices.

9 Conclusion

Our analysis began with a simple framework for understanding how selection patterns would respond to an attempt to decrease differential payments to MA plans via risk-adjustment. We predicted that individuals with the conditions included in the formula become more likely to join MA plans ("extensive-margin" selection falls), but their costs conditional on their risk score fall ("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. The framework also shows that because the variance of medical costs increases with the risk score, risk-adjustment can potentially increase the scope for selecting individuals with costs below their capitation payment, and we indeed find that differential payments to MA plans actually increase after risk-adjustment. We estimate that in 2006 they totaled \$30 billion, eight percent of total Medicare expenditures, or over one-fourth of the cost of extending insurance to over 30 million uninsured Americans through the Affordable Care Act.[45]

[44]To take an extreme example of how a very flexible risk-adjustment model can increase differential payments due to mismeasurement of the model's coefficients, suppose, for example, that CMS estimated a fully non-parametric version of the HCC model, where costs were estimated for every combination of the 70 HCC conditions. Such a model would have 2^{70} parameters, or trillions of times the number of FFS beneficiaries enrolled in a ten-year period. Here, many cells would have only one or two beneficiaries within them, and the corresponding coefficients would be measured with substantial error. Given the skewness of medical costs, some cells would be wildly over-priced, and going forward, MA plans would have strong incentives to attract individuals with these specific conditions, leading to large differential payments.

[45]The CBO estimates that in 2016 the total cost of the ACA's insurance expansion provisions will total \$114 billion. The CBO estimates the cost to be \$132 billion in 2016 dollars (see `http://www.cbo.gov/`

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 selection and differential payments change in response. The Medicare Prescription Drug Program (Part D) was established in 2006 and might interact with Medicare Advantage in interesting ways. Given that MA plans often provided drug coverage even before 2006, the extension of drug benefits to the entire Medicare population could change selection into the MA population. The Affordable Care Act more directly affects the Medicare Advantage program and aims to reduce differential payments by lowering many county benchmarks while at the same time linking capitation payments to measures of plan quality.[46]

While the focus of this paper has been on how imperfect risk adjustment affects the government's cost of providing health insurance, we close by considering several distributional issues. Given that, throughout our sample period, individuals enrolling in MA had lower costs than those remaining in FFS, our results suggest that this imperfect pricing not only causes the government to overpay for MA enrollees, but also shifts relative Medicare expenditure from high-cost to low-cost beneficiaries, relative to a system with perfect pricing or no MA option. It would thus seem to diminish Medicare's ability to smooth the utility consequences of variation in health status and thus lower the amount of social insurance per Medicare dollar spent.

Our results on satisfaction also suggest MA plans, relative to FFS, may devote more resources to lower-cost than higher-cost individuals. Both before and after risk adjustment, MA enrollees in poor health express greater dissatisfaction with their care than do their counterparts in FFS, and, not surprisingly, many return to FFS. Plans in the ACA insurance exchanges would seem to have a similar desire not to retain under-priced enrollees and thus might devote limited resources to their care, but these enrollees would not have a public, FFS-like option to which to return.[47] As such, we speculate that while at least some of the

ftpdocs/113xx/doc11379/AmendReconProp.pdf.) We use the CBO's forecasts for inflation to deflate this estimate to 2007 dollars, the units of the regression estimates.

[46] However, the extent to which MA reimbursements will actually fall remains an open question. The Department of Health and Human Services recently decided to award quality bonuses to MA plans covering the vast majority of MA beneficiaries, meaning that, by 2015, an additional $6.7 billion will flow to MA plans (see http://www.boston.com/business/healthcare/articles/2011/04/19/obama_administration_eases_pain_of_medicare_cuts/).

[47] "Optimal risk adjustment" models, pioneered by Glazer and McGuire (2000), have focused on settings without a public option—like the exchanges—and have shown that firms' incentives to skimp on the care of high-cost enrollees will lead any predictive risk-adjustment model calibrated on firm cost data to systematically under-estimate the cost of high-cost enrollees, further heightening the incentive to avoid such customers. For this reason they suggest that risk-adjustment needs to pay high- (low-) cost enrollees more (less) than their observed costs in firm data. Note that these models generally rely on the firm data being accurately reported to the risk adjustor, whereas, as we note, one might expect that firms would have the incentive to report inflated prices so as to increase capitation payments.

cost of imperfect pricing in the MA context appears to be born by the government and taxpayers in the form of higher Medicare costs, the under-priced consumers themselves may bear more of this cost in the exchanges.

These concerns suggest that governments may wish to take special care in "contracting out" social insurance. Imperfect pricing—whereby the government overpays a private firm relative to the cost and quality of in-house production—is, of course, a potential concern every time governments contract with a private party and has received great attention in the literature (see, for example, Hart *et al.* 1997). In the case of, say, paying a supplier for office equipment, the consequences of imperfect pricing would seem limited to whatever amount the government overpaid. With social insurance programs, however, imperfect pricing can induce strategic risk-selection, potentially reintroducing the very risk against which it was intended to insure. At least in the case of Medicare, we find little evidence that risk adjustment has solved this problem.

References

Aizer, A., Currie, J. and Moretti, E. (2007). Does managed care hurt health? Eevidence from Medicaid mothers. *The Review of Economics and Statistics*, **89** (3), 385–399.

Baicker, K. and Chandra, A. (2005). The effect of malpractice liability on the delivery of health care. *Forum for Health Economics & Policy*, **8**.

Batata, A. (2004). The effect of hmos on fee-for-service health care expenditures: evidence from medicare revisited. *Journal of Health Economics*, **23** (5), 951 – 963.

Bauhoff, S. (2010). Do health plans risk-select? an audit study on germanys social health insurance. *IRB*, **17130**, 101.

Blumenthal, D. and Glaser, J. (2007). Information technology comes to medicine. *New England Journal of Medicine*, **356** (24), 2527–2534.

Brown, J., Duggan, M., Kuziemko, I. and Woolston, W. (2010). *Can Risk Adjustment Reduce Selection in the Private Health Insurance Market? New Evidence from the Medicare Advantage Program*. Tech. Rep. 16977, NBER Working Paper.

Chandra, A., Jena, A. B. and Skinner, J. S. (2011). *The Pragmatists Guide to Comparative Effectiveness Research*. Working Paper 16990, National Bureau of Economic Research.

Chernew, M., DeCicca, P. and Town, R. (2008). Managed care and medical expenditures of medicare beneficiaries. *Journal of Health Economics*, **27** (6), 1451 – 1461.

Dafny, L. (2005). How do hospitals respond to price changes? *American Economic Review*, pp. 1525–1547.

de ven, W. V. and Ellis, R. (2000). Risk adjustment in competitive health plan markets. In A. Culyer and J. Newhouse (eds.), *Handbook of Health Economics*, vol. 1, Elsevier.

Duggan, M. (2004). Does contracting out increase the efficiency of government programs? evidence from medicaid hmos. *Journal of Public Economics*, **88** (12), 2549–2572.

Einav, L. and Finkelstein, A. (2011). Selection in insurance markets: Theory and empirics in pictures. *Journal of Economic Perspectives*, **25** (1), 115–38.

Ellis, R. and McGuire, T. (2007). Predictability and predictiveness in health care spending. *Journal of health economics*, **26** (1), 25–48.

Ellis, R. P. (2008). *Risk Adjustment in Health Care Markets: Concepts and Applications*, Wiley-VCH, pp. 177–222.

Fang, H., Keane, M. and Silverman, D. (2008). Sources of advantageous selection: Evidence from the medigap insurance market. *Journal of Political Economy*, **116** (2).

Finkelstein, A. and Poterba, J. (2006). Testing for adverse selection with "unused observables".

Frogner, B., Anderson, G., Cohen, R. and Abrams, C. (2011). Incorporating new research into medicare risk adjustment. *Medical Care*, **49** (3), 295.

Glazer, J. and McGuire, T. G. (2000). Optimal risk adjustment in markets with adverse selection: An application to managed care. *American Economic Review*, **90** (4), 1055–1071.

Hall, A. (2011). Measuring the return on government spending on the medicare managed care program. *The BE Journal of Economic Analysis & Policy*, **11** (2), 7.

Hart, O., Shleifer, A. and Vishny, R. (1997). The proper scope of government: Theory and an application to prisons*. *Quarterly Journal of Economics*, **112** (4), 1127–1161.

Holmstrom, B. and Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization*, **7** (0), 24–52.

Langwell, K. and Hadley, J. (1989). Evaluation of the Medicare competition demonstrations. *Health Care Financing Review*, **11** (2), 65–80.

Lazear, E. P. (2006). Speeding, terrorism, and teaching to the test. *The Quarterly Journal of Economics*, **121** (3), 1029–1061.

Lumley, T., Diehr, P., Emerson, S. and Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, **23** (1), 151–169.

Lustig, J. (2010). Measuring welfare losses from adverse selection and imperfect competition in privatized medicare. *Manuscript. Boston University Department of Economics.*

Mello, M. M., Stearns, S. C., Norton, E. C. and Ricketts, T. C. (2003). Understanding biased selection in medicare hmos. *Health Services Research*, **38**, 961–992(32).

Newhouse, J., Buntin, M. and Chapman, J. (1997). Risk adjustment and medicare: taking a closer look. *Health Affairs*, **16** (5), 26.

Nicholas, L. (2009). *Does Managed Care Reduce Federal Spending? Evidence from Medicare*. Tech. Rep. 09-671, Population Studies Center Research Report.

Physician Payment Review Commission (1997). Testimony of gail r. wilensky. statement before the us house of representatives subcommittee on health. committee on ways and means, february 24, 1997.

Pope, G., Kautter, J., Ellis, R., Ash, A., Ayanian, J., Iezzoni, L., Ingber, M., Levy, J. and Robst, J. (2000). Diagnostic cost group hierarchical condition category models for medicare risk adjustment: Final report prepared for health care financing ad-

ministration.

Pope, G. C., Kautter, J., Ellis, R. P., Ash, A. S., Ayanian, J. Z., Iezzoni, L. I., Ingber, M. J., Levy, J. M. and Robst, J. (2004). Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financing Review*, **25** (4), 119–141.

Saltman, R. and Figueras, J. (1998). Analyzing the evidence on european health care reforms. *Health Affairs*, **17** (2), 85–108.

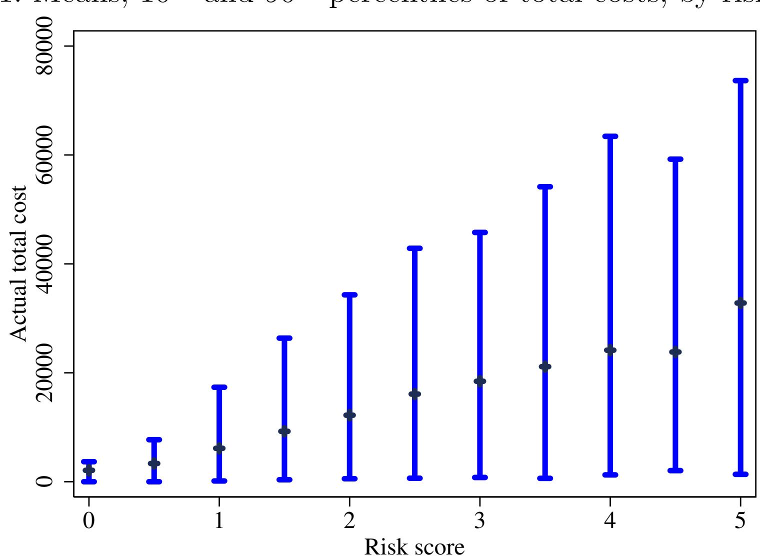
Shleifer, A. (1985). A theory of yardstick competition. *The RAND Journal of Economics*, **16** (3), 319–327.

Silverman, E. and Skinner, J. (2004). Medicare upcoding and hospital ownership. *Journal of Health Economics*, **23** (2), 369–389.

Town, R. and Liu, S. (2003). The welfare impact of medicare hmos. *The RAND Journal of Economics*, **34** (4), 719–736.

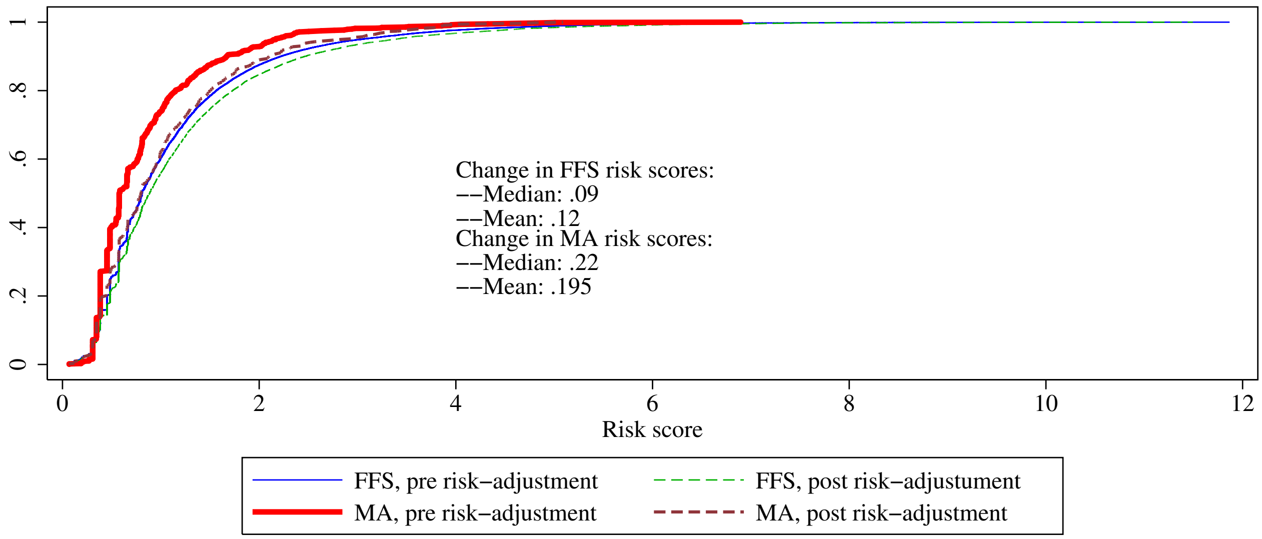
Yabroff, K. R., Lamont, E. B., Mariotto, A., Warren, J. L., Topor, M., Meekins, A. and Brown, M. L. (2008). Cost of care for elderly cancer patients in the united states. *Journal of the National Cancer Institute*, **100**, 630–641(12).

Figure 1: Means, 10^{th} and 90^{th} percentiles of total costs, by risk score



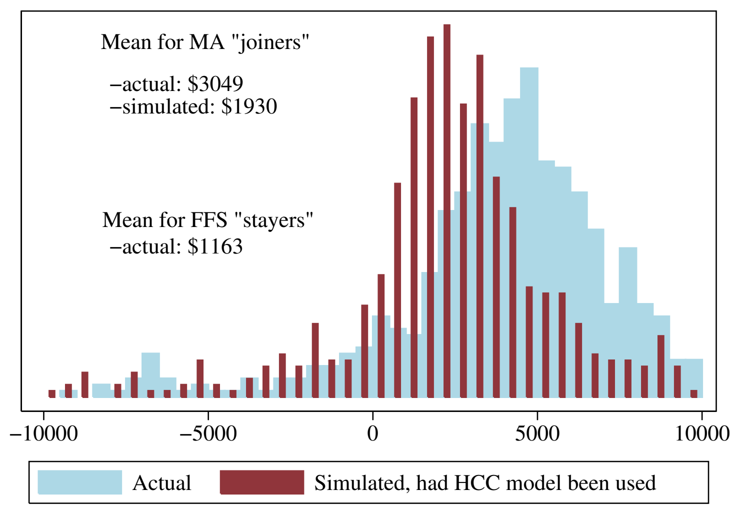
Notes: All observations spent all twelve months of the previous year in FFS (so that current-year risk scores can be calculate) and no months of the current year in MA (so that all current cost data can be observed). Observations are taken only from the pre-period so that the sample is unlikely to be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

Figure 2: Cumulative risk score distribution for individuals switching to MA



Notes: All observations spent all twelve months of the previous year in FFS (so that a full year of claims data is recorded, which we use to calculate the risk score). Individuals are counted as being in MA if they spent at least half their Medicare-eligible months of the baseline year in MA and are counted as being in FFS others. Dollar amounts are adjusted to 2007 dollars using the CPI-U. Sample weights provided by the MCBS are used.

Figure 3: Medicare expenditure increases for enrollees switching from FFS to MA, 1994 to 2003



Notes: The two histograms are based on observations that spend at least healf of their Medicare-eligible months in the baseline year in MA and spent all twelve months of the previous year in FFS (so that a full year of claims data is recorded, which we use to calculate the risk score). The statistics for FFS "stayers" are based on observations that spend the majority of their Medicare-eligible months in the baseline year in FFS and also spent all twelve months of the previous year in FFS. Dollar amounts are adjusted to 2007 dollars using the CPI-U. Sample weights provided by the MCBS are used.

Table 1: Changes in selection patterns after risk adjustment

	Extensive-margin selection					Intensive-margin selection			
	Dependent variable: Risk score					Dependent var: Total Medicare expenditure			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Fraction of next year spent in MA	-0.305***	-0.180***	-0.233***		-0.238***	171.5		191.6	-2847.0***
	[0.0355]	[0.0272]	[0.0325]		[0.0433]	[316.5]		[467.2]	[396.9]
Fraction of next year in MA x After 2002	0.106*	0.116**	0.140***		0.145**	-1217.9**		-1280.3*	-172.7
	[0.0614]	[0.0509]	[0.0507]		[0.0568]	[604.0]		[691.8]	[713.4]
In MA majority of next year				-0.176***			229.4		
				[0.0272]			[253.4]		
In MA majority of next year x After 2002				0.0827*			-1045.3**		
				[0.0437]			[511.3]		
HCC score						9903.4***	9903.6***	9691.0***	
						[182.5]	[182.5]	[202.2]	
Mean of dept. var.	1.185	1.008	1.185	1.185	1.211	6491.9	6491.9	6630.7	6491.9
Estim. method	OLS	OLS	Q. reg.	Q. reg.	Q. reg.	OLS	OLS	OLS	OLS
Outliers trimmed	No	Yes	No	No	No	No	No	No	No
1997-2005 only	No	No	No	No	Yes	No	No	Yes	No
Observations	73,054	69,266	73,054	73,054	54,646	73,054	73,054	54,646	73,054

Notes: All observations are in FFS all twelve months of the given year. Year fixed effects included in all regressions. The outcome in cols. (1) through (5) is an individual's HCC score the following year, which is based on current-year claims. The outcome in cols. (6) through (9) is an individual's current year total Medicare expenditure. "Q. reg" refer to median regressions. "Outliers trimmed" excludes individuals with risk scores above the 95$^{\text{th}}$ percentile (where percentiles are calculated separately by year). 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. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Table 2: Changes in differential payments after risk adjustment

	Dependent variable: Total Medicare expenditure								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Share of months in MA x After 2003	1733.4^{**}	2268.7^{***}	2224.1^{***}	2090.3^{***}		1843.2^{**}	2010.6^{**}	3257.6^{***}	2778.6^{***}
	[747.0]	[789.3]	[776.0]	[810.9]		[747.0]	[882.0]	[828.1]	[747.0]
Share of months in MA	905.2^{***}	624.6^{**}	672.2^{**}	768.0^{**}		1199.3^{***}	870.6^{*}	767.5^{**}	671.8^{*}
	[256.5]	[286.9]	[286.3]	[349.7]		[308.4]	[495.1]	[349.9]	[352.2]
In MA majority of year					729.7^{**}				
					[289.9]				
In MA majority of year x After 2003					1554.4^{**}				
					[686.5]				
Benchmarks adjusted	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Baseline controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged health controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Dept. var windsorized	No	No	No	No	No	Yes	No	No	No
Only 1998-2006	No	No	No	No	No	No	Yes	No	No
MCBS cap. payments	No	No	No	No	No	No	No	No	Yes
Observations	73,054	72,930	72,638	72,375	72,375	72,375	54,120	72,375	72,375

Notes: All observations are in FFS all twelve months of the previous year. Year fixed effects are included in all regressions, and county fixed effects included in col. (2) - (9). All regressions include a once-lagged dependent variable, as well as dummy variables corresponding to eleven bins of lagged Part A and B expenditure (with zero as its own bin and ten bins corresponding to ten deciles of positive Part A and B expenditure, calculated separately for each year). "Benchmarks adjusted" refers to reducing MA payments after 2003 in the following manner: the growth rate in county benchmarks is constrained to match that of the pre-period, and the "budget neutrality" adjustment meant to ease the risk-adjustment process is eliminated. Both of these adjustments make it less likely that the interaction term would have a positive coefficient, as one can see from comparing cols. (4) and (8). "MCBS cap. payments" refers to using the uncorrected capitation payments in the MCBS as a dependent variable. "Baseline controls" include the following: individual's predicted capitation payment based on the demographic model; race and Hispanic origin; gender; age-in-year fixed effects; fixed effects for eligibility status (disabled and old-age, with and without end-stage-renal disease as a secondary condition); Medicaid status; the interaction of disability status and Medicaid status; income category fixed effects; months of Medicare eligibility; and education category fixed effects. "Lagged health controls" includes fixed effects for the five categories of lagged self reported health (excellent, very good, good, fair, poor), the lagged share of the year spent in an institution, and the lagged risk score. "Health controls" include the following: five categories of current self-reported health, the difference between current and previous-year self-reported health, an indicator variable for being alive the entire year, and the share of the year spent in an institution. The dependent variable is windsorized at the 99^{th} percentile in col. (6). 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. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 3: Effect of MA enrollment and health status on enrollee satisfaction

Dependent var:		OLS coefficient estimates (clustered SEs)	
Satisfaction rating (1-4)	Obs.	In MA	MA x Health (demeaned)
Overall medical care	75,886	-0.0180**	0.0141**
		(0.00748)	(0.00667)
Out-of-pocket costs	75,311	0.0556***	0.0226***
		(0.00868)	(0.00748)
Follow-up care	69,766	0.00428	0.00611
		(0.00676)	(0.00608)
Doctor's concern for your health	74,713	-0.0123*	0.0153**
		(0.00717)	(0.00637)
Information about your medical condition	75,542	-0.00163	0.0128**
		(0.00676)	(0.00596)
Access to specialists	57,189	-0.0212***	0.000524
		(0.00745)	(0.00657)
Questions answered over phone	48,619	-0.0127	0.0195**
		(0.00891)	(0.00780)
Availability of care nights and weekends	44,503	0.0131	0.00832
		(0.00880)	(0.00803)
Medicare care provided in same location	69,382	0.0442***	0.00907
		(0.00664)	(0.00581)
Retains coverage type next year (probit coefficients)	82,145	-0.545***	0.0787***
		(0.0274)	(0.0213)

Notes: Each row represents a regression of the form: $satisfaction\, category_i = \beta_1 MA_i + \beta_2 MA_i \times Health_i + \gamma \mathbf{H}_i + \lambda \mathbf{X}_i + \epsilon_i$, 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 give year, $Health$ is a (demeaned) linear measure of the five-category self-reported health variable, $\mathbf{H}$ is a vector of fixed effect for the five health categories (one, "poor," up to five, "excellent"), and $\mathbf{X}$ is a vector of basic controls: age, state-of-residence, and year fixed effects, and indicator variables for being female, disabled, or on Medicaid. 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 $MA \times Health$ 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. Note that this table is based on all observations in the survey, instead of just those who are in the survey at least two years, the subsample used in much of the previous analysis. Sample weights provided by the MCBS are used and standard errors are clustered by individual. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendices for online publication only

Appendix Table 1: The ten most common conditions in the HCC formula

Category	Prevalence	Description	HCC weight
80	0.126	Congestive Heart Failure	0.417
108	0.124	Chronic Obstructive Pulmonary Disease	0.376
19	0.120	Diabetes without Complication	0.200
92	0.097	Specified Heart Arrhythmias	0.266
105	0.094	Vascular Disease	0.357
10	0.063	Breast, Prostate, Colorectal Cancers	0.233
83	0.046	Angina Pectoris/Old Myocardial Infarction	0.235
96	0.045	Ischemic or Unspecified Stroke	0.306
38	0.039	Rheum. Arthritis and Inflam. Connective Tissue Disease	0.322
79	0.038	Cardio-Respiratory Failure and Shock	0.692

Notes: This table is based on the FFS population, 1993-2006. The weight associated with each HCC condition is added to a person's total risk score. Given that the average benchmark is roughly $9,345—average per capita FFS expenditure ($8,344) multiplied by the benchmark-to-FFS markup in 2006 (1.12)—in 2006, having been diagnosed with congestive heart failure in the previous year would mean an individual's capitation payment is increased by $0.417 * \$9,345 = \$3,897$.

Appendix Table 2: Summary statistics, Medicare Current Beneficiary Survey 1994-2006

	FFS	MA	Difference: FFS - MA
On Medicaid	0.16	0.070	0.092***
Disabled	0.13	0.067	0.061***
Age	72.8	74.0	-1.12***
Age, those not disabled	76.2	75.4	0.88***
Female	0.57	0.57	-0.0012
Black	0.092	0.098	-0.0057**
temp_hisp	0.018	0.034	-0.015***
Resides in a metro area	0.71	0.97	-0.25***
Self-reported health	3.12	3.35	-0.23***
Income > 20,000	0.45	0.48	-0.028***
Has a high school degree	0.75	0.74	0.0035
Total Medicare expenditure (risk scores)	7304.7	7870.9	-566.2***
Total Medicare expenditure	7310.1	8044.8	-734.7***
Observations	131,339	19,320	

Notes: All observations taken from 1994-2006 MCBS. An individual is classified as in FFS if he spends the majority of his Medicare-eligible months in FFS and as in MA otherwise. Self-reported health varies from one (poor) to five (excellent). *Total Medicare expenditure* includes any capitation payments and Part A and B payments made by Medicare on behalf of an individual in a given year. Note that this table is based on all observations in the survey, instead of just those who are in the survey at least two years, a subsample used in much of the empirical analysis. Sample weights provided by the MCBS are used. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Appendix Table 3: 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) $\rightarrow$ FFS (year t+1)	19,017	18,539	18,305	17,329	73,190
FFS (year t) $\rightarrow$ MA (year t+1)	566	399	102	464	1,531
MA (year t) $\rightarrow$ FFS (year$t$+1)	102	165	457	125	849
MA (year t) $\rightarrow$ 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

Notes: 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.

Appendix Table 4: Summarizing changes in incentives after risk adjustment

HCC score	HCC payment minus demographic payment	HCC payment minus actual Medicare expenditure
0-25$^{\text{th}}$ percentile (lowest scores)	-2,993	67
25-50$^{\text{th}}$ percentile	-2,406	198
50-75$^{\text{th}}$ percentile	-342	549
75-99$^{\text{th}}$ percentile	6,701	893
99-100$^{\text{th}}$ 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 twelve 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.

Appendix Table 5: Changes in differential payments after risk adjustment, using simulated instead of administrative risk scores

	Dependent variable: Total Medicare expenditure								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Share of months in MA x After 2003	1707.1**	2242.4***	2211.8***	2078.0**		1815.9**	1996.2**	3237.5***	2778.6***
	[750.4]	[796.1]	[781.7]	[815.4]		[744.7]	[886.0]	[832.1]	[747.0]
Share of months in MA	906.6***	626.7**	674.7**	770.4**		1201.7***	874.9*	770.4**	671.8*
	[256.5]	[286.9]	[286.3]	[349.7]		[308.4]	[495.1]	[349.8]	[352.2]
In MA majority of year					731.5**				
					[289.9]				
In MA majority of year x After 2003					1547.4**				
					[689.9]				
Benchmarks adjusted	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Baseline controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged health controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Dept. var windsorized	No	No	No	No	No	Yes	No	No	No
Only 1998-2006	No	No	No	No	No	No	Yes	No	No
MCBS cap. payments	No	No	No	No	No	No	No	No	Yes
Observations	73,054	72,930	72,638	72,375	72,375	72,375	54,120	72,375	72,375

Notes: All observations are in FFS all twelve months of the previous year. Year fixed effects are included in all regressions, and county fixed effects included in col. (2) - (9). All regressions include a once-lagged dependent variable, as well as dummy variables corresponding to eleven bins of lagged Part A and B expenditure (with zero as its own bin and ten bins corresponding to ten deciles of positive Part A and B expenditure, calculated separately for each year). "Benchmarks adjusted" refers to reducing MA payments after 2003 in the following manner: the growth rate in county benchmarks is constrained to match that of the pre-period, and the "budget neutrality" adjustment meant to ease the risk-adjustment process is eliminated. Both of these adjustments make it less likely that the interaction term would have a positive coefficient, as one can see from comparing cols. (4) and (8). "MCBS cap. payments" refers to using the uncorrected capitation payments in the MCBS as a dependent variable. "Baseline controls" include the following: individual's predicted capitation payment based on the demographic model; race and Hispanic origin; gender; age-in-year fixed effects; fixed effects for eligibility status (disabled and old-age, with and without end-stage-renal disease as a secondary condition); Medicaid status; the interaction of disability status and Medicaid status; income category fixed effects; months of Medicare eligibility; and education category fixed effects. "Lagged health controls" includes fixed effects for the five categories of lagged self reported health (excellent, very good, good, fair, poor), the lagged share of the year spent in an institution, and the lagged risk score. "Health controls" include the following: five categories of current self-reported health, the difference between current and previous-year self-reported health, an indicator variable for being alive the entire year, and the share of the year spent in an institution. The dependent variable is windsorized at the 99$^{\text{th}}$ percentile in col. (6). 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. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

10 Theoretical Framework Appendix

In this Appendix, we formalize the intuition provided in Section 3. The purpose of this model is to understand how adopting risk adjustment will influence total costs to the government from offering MA plans. We therefore take as given the basic contours of the risk-adjustment formula used by CMS, as opposed to exploring the optimal formula, as in Glazer and McGuire (2000) and others.

While an MA plan must be open at the same price to all individuals in the plan's geographic area of operation, the model assumes that, as shown in earlier work, plans have at least *some* scope to encourage individual with certain characteristics to enroll. For example, by differentially advertising in *Diabetes Forecast* (a publication of the American Diabetes Association), MA plans could increase the probability that diabetics enroll.

We emphasize that this process does not necessarily imply that the firm have access to information about the characteristics of any individual Medicare beneficiaries. Instead, firms could use information on the conditional distribution of costs in the Medicare population and employ strategies, such as targeted advertising or changing the quality of physicians in their network, to encourage beneficiaries with certain conditions to enroll. Beneficiaries, who have private information on their health type, choose to enroll in MA based on the perceived costs and benefits of the plan.[48]

To keep the model tractable, we do not model the consumer side of the enrollment decision and instead focus on firms' decision to incur the costs associated with these screening activities in return for enrolling a selected subsample from the Medicare population. In our model, firms have an incentive to target individuals for whom the difference between capitation payments and expected costs is the greatest, and risk-adjustment changes this set of individuals by changing how capitation payments are calculated.

10.1 Basic framework and assumptions

10.1.1 Cost of health insurance coverage

Let the cost of covering individual i in a given year be given by $m_i = b_i + v_i$, where b_i is an individual's expected cost conditional on the variables included in the risk-adjustment formula used by the government, and v_i is the residual. As MA contracts have a year-long duration, the model is single-period, and we thus specify costs over a single year.[49] Both v and b are in units of absolute dollars.[50] While $\mathbb{E}(v|\, b) = 0$ for all b, the conditional variance of v can vary with b, consistent with past work showing substantial heteroskedasticity in medical costs. We assume that costs m are the same whether an individual is in FFS or MA.

[48]Note that the model does not rule out the possibility that firms use some information to actively encourage some individuals to enroll in their plan. For example, MA plans may respond more quickly to enrollment requests from respondents residing in low-cost areas, as Bauhoff (2010) finds in the German context.

[49]We return to the question of dynamics in Section 8 when we discuss recalibrating the risk-adjustment model over time.

[50]Note that m is the cost to the *insurer*—the cost of total medical care plus administrative costs, less the out-of-pocket costs paid by the individual—not total actual medical costs. As in Glazer and McGuire (2000), we do not model out-of-pocket costs in order to focus on selection, though we present results on individuals' satisfaction with their out-of-pocket costs in Section 7.

Of course, MA plans may be better or worse at controlling costs than FFS, and all of the results that follow hold when MA costs are proportional to FFS costs. However, we focus on the case where costs are identical. This assumption not only simplifies the analysis, but also allows us to more easily focus on the difference between payments to private firms for insuring person i and the counterfactual cost if the government directly covered her, which is a key parameter for evaluating the fiscal impact of private Medicare Advantage plans.[51]

10.1.2 Capitation payments and risk adjustment

Without risk-adjustment, firms receive a fixed payment $\bar{p}$ for each individual they enroll. We model risk-adjustment as replacing $\bar{p}$ with a function $p(b)$, $p' > 0$, so that capitation payments become an increasing function of b. While our main results on selection and differential payments do not require that risk- adjusted payments are linear in b, this assumption corresponds to the MA setting where capitation payments are calculated by multiplying risk scores by a fixed county factor. As it allows us to generate additional empirical predictions and also simplifies the analysis, we take as a baseline assumption that $p''(\cdot) = 0$.[52]

We also make risk-adjustment be "payment-neutral," that is, $\mathbb{E}(p(b)) = \bar{p}$ for the Medicare population as a whole. In other words, if the entire population joined a private plan, the government would pay the same average capitation payment with or without risk adjustment.[53]

Finally, we want to allow for the degree of risk-adjustment to vary, which again mirrors the actual experience of the phasing-in of risk adjustment between 2004 and 2007. We define capitation payments as $(1-\Omega)\bar{p} + \Omega p(b_i)$, where $\Omega \in [0, 1]$ is the risk-adjusted share of the capitation payment.

As indicated in the introduction, the key objective of risk adjustment was to reduce the difference between a plan's capitation payment for covering an individual and the cost to the government had it directly covered him via FFS. Having defined how risk-adjustment affects capitation payments, we can make this concept slightly more precise.

Definition. *The "differential payment" for individual i equals*

$$\underbrace{(1-\Omega)\bar{p} + \Omega p(b_i)}_{\text{capitation payment}} - \underbrace{(b_i + v_i)}_{\text{FFS cost}}$$

.

[51] Whether the HMO model is actually more efficient than the fee-for-service model even absent selection effects is an open question. Duggan (2004) finds that when some California counties mandated their Medicaid recipients to switch from the traditional FFS system to an HMO, costs increased by 17 percent relative to counties that retained FFS. As, within a county, individuals did not select between FFS or an HMO, selection issues are unlikely to be driving the result.

[52] In particular, our proofs of Proposition 1 (that risk adjustment causes selection to fall along the b margin and rise along the v margin) and Proposition 3 (that the effect of risk-adjustment on differential payments is ambiguous) do not depend on the linearity of $p(\cdot)$.

[53] As we discuss in Section 2, firms were actually given temporary payments to ease the transition into risk-adjustment, but as a matter of theory, we are more interested in the steady-state results when the system returns to payment-neutral conditions. Section 6 reports our empirical results with and without these temporary payments.

10.1.3 Screening costs

Though we discuss profit-maximization in greater detail shortly, firm profits are obviously a function of an individual's cost $m_i = b_i + v_i$, and thus firms will have preferences over the b and v values of their enrollees, even if the firm is unable to observe b and v for any potential beneficiary. However, MA plans are required to accept any patient in their geographic coverage area who chooses to enroll, and selectively encouraging certain individuals to enroll will entail screening costs. Thus, even though firms cannot directly control the characteristics of their beneficiaries, because firms can indirectly influence the population who signs up, we assume that b and v are choice variables on the part of the firm.

We assume that the per capita screening cost c a firm incurs is given by $c(b, v)$, where b and v are its enrollees' average values of b_i and v_i. Since randomly enrolling individuals from the general population should require minimum screening costs, $c(\bar{b}, \bar{v})$ is a global minimum, where $\bar{b}$ and $\bar{v}$ are population averages (recall we assume $\bar{v} = 0$). Encouraging individuals to enroll who are further from the mean is costly, so $c_x < 0$ for $x < \bar{x}$ and $c_x > 0$ for $x > \bar{x}$ for $x \in \{b, v\}$. We also assume that the cost function is everywhere convex.

Finally, we assume that $c_{bv} > 0$. This assumption implies that for higher values of b, the incremental cost of reducing v falls. This assumption rules out the possibility that screening in b and v are complements. Because the variance of medical costs is typically a positive function of expected costs (see, e.g., Lumley *et al.* 2002 and Figure XX) and v is measured in absolute dollars, it should be easier to attract, say, a cancer patient with costs \$100 below what her risk score would predict than someone without a single documented disease condition with costs \$100 below what her risk score would predict.

With screening costs thus defined, we can now specify a firm's profit function. In our baseline model, we make the simplifying assumption that firms cannot affect the number of individuals that they enroll, though we return to this assumption later in the section. Firms instead focus on maximizing the average profit per enrollee, which is a function of b and v. Thus, firms maximize the following expression:

$$\mathbb{E}(\pi) = \underbrace{(1-\Omega)\bar{p} + \Omega p(b)}_{\text{capitation payment}} - \left(\underbrace{b + v}_{\text{FFS cost}} \right) - \underbrace{c(b, v)}_{\text{screening cost}} \quad . \tag{5}$$

We now use this framework to prove a number of results regarding selection and differential payments.

10.2 Results

We begin with our main selection result, which characterizes how firms will react to a change in risk adjustment.

Proposition 1. *The following two conditions hold when the risk-adjusted share Ω of the capitation payment increases:*

(i) Firms decrease screening along the b margin and thus the average value of b among their enrollees rises ("extensive-margin" selection decreases).

(*ii*) *Firms increase screening along the v margin and thus the average value of v among their enrollees falls ("intensive-margin" selection increases).*

This proposition formalizes the result from the Theoretical Framework that (1) "the risk scores of those enrolling in MA will increase relative to those remaining in FFS" (2) "actual costs conditional on the risk score will fall among those enrolling in MA relative to those remaining in FFS."

Proof. We are required to show that $\frac{\partial b^*}{\partial \Omega} > 0$ and $\frac{\partial v^*}{\partial \Omega} < 0$, where b^* and v^* are a firm's optimal levels of b and v. The first-order conditions from maximizing the profit expression in equation (5) with respect to b and v are given by

$$[b] \quad : \quad \Omega p'(b^*) - c_b(b^*, v^*) = 1 \tag{6}$$

$$[v] \quad : \quad -c_v(b^*, v^*) = 1 \tag{7}$$

Totally differentiating equation (6) with respect to Ω yields

$$p'(\cdot) + \Omega p''(\cdot)\frac{\partial b^*}{\partial \Omega} - c_{11}(\cdot,\cdot)\frac{\partial b^*}{\partial \Omega} - c_{12}(\cdot,\cdot)\frac{\partial v^*}{\partial \Omega} = 0 \tag{8}$$

Similarly, equation (7) yields:

$$c_{bv}(\cdot,\cdot)\frac{\partial b^*}{\partial \Omega} + c_{vv}(\cdot,\cdot)\frac{\partial v^*}{\partial \Omega} = 0$$

or

$$\frac{\partial v^*}{\partial \Omega} = -\frac{c_{bv}}{c_{vv}}\frac{\partial b^*}{\partial \Omega}. \tag{9}$$

Substituting equation (9) into (8) gives:

$$p'(\cdot) + \Omega p''(.)\frac{\partial b^*}{\partial \Omega} - c_{bb}(\cdot,\cdot)\frac{\partial b^*}{\partial \Omega} - c_{bv}(\cdot,\cdot)\left(-\frac{c_{bv}}{c_{vv}}\frac{\partial b^*}{\partial \Omega}\right) = 0.$$

We can now solve for $\frac{\partial b^*}{\partial \Omega}$ and sign many of the terms:

$$\frac{\partial b^*}{\partial \Omega} = \frac{\overbrace{p'}^{+\text{ as cap payments increase in } b}}{-\Omega p''(\cdot) + \frac{\overbrace{(c_{bb}c_{vv}) - c_{bv}^2}^{+\text{ by convexity of } c(\cdot,\cdot)}}{\underbrace{c_{vv}}_{+\text{ by convexity of } c(\cdot,\cdot)}}} \tag{10}$$

By assumption, $p''(\cdot) = 0$, so the entire denominator is positive. As $\frac{\partial b^*}{\partial \Omega} > 0$ and $c_{bv}, c_{vv} > 0$, equation (9) gives the result in (ii). ∎

As risk-adjustment makes capitation payments a positive function of b, firms will spend

less effort finding low-b enrollees and instead focus on finding low-v enrollees. We term the first result "extensive-margin" selection as it relates to the government's risk score, which is an approximate measure of actual cost; we term the second result "intensive-margin" selection because it relates to how intensely individuals are selected conditional on the risk score.[54]

Proposition 2. *For $\Omega_0 < \Omega_1$, moving from Ω_0 to Ω_1 will always decrease differential payments if (1) b and v are held fixed at their equilibrium values under Ω_0 and (2) if individuals are positively selected with respect to b under Ω_0.*

This proposition formalizes the result from the Theoretical Framework that, "applying the risk-adjustment formula to the pre-risk-adjustment population of MA enrollees would have decreased the total capitation payments the government would have made on their behalf."

Proof. The result is easy to show when $p(\cdot)$ is linear. Recall that $p(\cdot)$ is "payment-neutral," so that $\mathbb{E}(p(b)) = \bar{p}$. For linear p, $\mathbb{E}(p(b)) = p(\bar{b}) = \bar{p}$, so risk-adjustment does not change the payment for an individual with $b = \bar{b}$. As $p' > 0$, $p(b) < p(\bar{b}) = \bar{p}$ for all $b < \bar{b}$. So, as long as individuals are positively selected with respect to b under Ω_0 ($b < \bar{b}$), the proposition holds. ■

Proposition 3. *The effect of increasing Ω on a firm's average differential payment is ambiguous.*

Proof. Let $\phi(\Omega)$ denote the differential payment when the risk-adjusted share of the capitation payment is set to Ω and firms are at their optimal b and v values:

$$\phi(\Omega) = \underbrace{\Omega p(b^*(\Omega)) + (1-\Omega)\bar{p}}_{\text{capitation payment}} - \underbrace{(b^*(\Omega) + v^*(\Omega))}_{\text{actual costs}} \tag{11}$$

Differentiating with respect to Ω gives:

$$\phi'(\Omega) = \Omega p' \frac{\partial b^*}{\partial \Omega} + p(b^*) - \bar{p} - \frac{\partial b^*}{\partial \Omega} - \frac{\partial v^*}{\partial \Omega} \tag{12}$$

Rearranging and substituting $\frac{\partial v^*}{\partial \Omega} = -\frac{c_{bv}}{c_{vv}} \frac{\partial b^*}{\partial \Omega}$ from equation (9) yields

$$\phi'(\Omega) = [p(b^*) - \bar{p}] + \frac{\partial b^*}{\partial \Omega}(\Omega p' - 1 + \frac{c_{bv}}{c_{vv}}) \tag{13}$$

We showed in the proof of Proposition 2 that $p(b^*) < \bar{p}$ for any equilibrium b^*, so the first term (in brackets) is negative. However, the second term is ambiguous. While Ω and

[54] The empirical work will focus on the government's observed risk score—that is, $p(b)$ in the parlance of the model—as b itself is not observable. But as $p'(b) > 0$, Proposition (1) (i) implies that $p(b)$ will increase as well, thus giving the testable prediction that risk scores *as measured by the government* increase with an increase in risk-adjustment.

p' are both by assumption less than one and $\frac{\partial b^*}{\partial \Omega} > 0$ by Proposition 1, if $\frac{c_{bv}}{c_{vv}}$ is large, the expression can indeed be positive. This condition requires c_{bv} to be sufficiently positive. ■

Endogenizing firm enrollment size

We now assume that firms maximize *total*, as opposed to per capita, profits which equal $q(b,v)\pi(b,v,\Omega)$, where π is average per capita profits as specified in equation (5) and q is the number of enrollees the firm has.

The first-order conditions with respect to b and v are now:

$$[b] \quad : \quad q_b(b,v)\pi(b,v,\Omega) + q(b,v)\underbrace{(\Omega p' - 1 - c_b(b,v))}_{\pi_b} \tag{14}$$

$$[v] \quad : \quad q_v(b,v)\pi(b,v,\Omega) + q(b,v)\underbrace{(-1 - c_v(b,v))}_{\pi_v} \tag{15}$$

Note that when the level of q is larger relative to (i) its partial derivatives or (ii) the level of per capita profits, then equations (14) and (15) reduce to the original first-order conditions of $\pi_b = \pi_v = 0$.