

# Upcoding or Selection?

## Evidence from Medicare on Squishy Risk Adjustment\*

Michael Geruso<sup>†</sup>      Timothy Layton<sup>‡</sup>

October 13, 2014

### Abstract

Risk adjustment is commonly used in health insurance markets to deal with problems of adverse selection and cream skinning by compensating health plans for insuring consumers whose diagnoses imply high expected costs. However, in all real world risk adjustment systems, it is the insurers themselves who report the diagnoses that determine risk scores. This creates incentives to “upcode” enrollees to extract higher payments. We model upcoding in the presence of adverse selection. Our model delivers a novel strategy for empirically separating upcoding from selection in aggregate, market-level data. We apply this strategy to analyze upcoding by Medicare Advantage plans. The results show that enrollees in Medicare Advantage plans generate 7% higher risk scores on average (and therefore 7% higher payments) than what the same enrollees would generate under Traditional Medicare. Absent a coding inflation correction, this implies a distortion in seniors’ choice between Medicare Advantage and Traditional Medicare, and excess payments to Medicare Advantage of \$11.4 billion annually. We find that the degree of upcoding increases in the level of insurer-provider integration, suggesting that more integrated plans are better able to align incentives for their providers to code intensively.

---

\*We thank seminar participants at the 2014 American Society of Health Economists Biennial Meeting, Southeastern Health Economics Study Group, RTI, Harvard Medical School, the University of Texas at Austin, Boston University, and the University of Illinois at Chicago for useful comments. We also thank Chris Afendulis, Francesco Decarolis, Randy Ellis, Tom McGuire, Joe Newhouse, and Daria Pelech for useful conversations. We thank Joshua Gottlieb and Colleen Carey for serving as discussants. We gratefully acknowledge financial support from the National Institute of Mental Health (Layton, T32-019733).

<sup>†</sup>University of Texas at Austin and NBER. Email: [mike.geruso@gmail.com](mailto:mike.geruso@gmail.com)

<sup>‡</sup>Harvard Medical School. Email: [timothyjlayton@gmail.com](mailto:timothyjlayton@gmail.com)

# 1 Introduction

Risk adjustment is widely used in health insurance markets to counteract inefficient selection. By modifying payments to insurers to compensate for enrolling a high expected cost consumer, risk adjustment weakens incentives for insurers to distort product characteristics toward attracting lower-cost enrollees, as in [Rothschild and Stiglitz \(1976\)](#).<sup>1</sup> In a setting where product characteristics are fixed, risk adjustment also breaks the link between a plan's price and the average cost of its enrollees, which could otherwise cause inefficient sorting or market unravelling ([Akerlof, 1971](#); [Einav, Finkelstein and Cullen, 2010](#)). While the intuition underlying risk adjustment is straightforward, the mechanism relies on an implicit assumption that the regulator can construct consumer "risk scores" that summarize consumer health risk in a way that is invariant to the insurer with whom they are enrolled. But in all real world payment systems, it is the insurer itself that reports the diagnoses that determine enrollee risk scores and ultimately the insurer's payments. Whether and to what extent risk scores can be influenced by insurers is of significant practical importance, as risk adjustment is the primary regulatory mechanism used to counteract distortions caused by asymmetric information in US health insurance markets.<sup>2</sup>

In this paper, we show that even when successful in counteracting cream-skimming and inefficient sorting due to adverse selection, risk adjustment introduces a new distortion: It incentivizes intensive coding by insurers. We begin with a simple theoretical framework that illuminates how coding intensity differences across insurers or across market segments can impact total spending on publicly funded insurance programs like Medicare and Medicaid. We also characterize the consumer choice distortions that arise from the implicitly larger regulatory transfer or subsidy to plans that practice higher coding intensity. We then provide the first econometrically identified estimates of coding intensity differences, here analyzing coding differences between the uncoordinated reporting of diagnoses in the Traditional Medicare system (TM) and Medicare Advantage (MA), the private Medicare option in which premiums are heavily or fully subsidized by risk-adjusted capitation payments to private insurers.

---

<sup>1</sup>For instance in Medicare Advantage, a diagnosis for the condition Diabetes with Chronic Complications generates a payment for the private insurer who enrolls the patient that is incrementally larger by about \$3,665 per year, the average incremental cost incurred by individuals diagnosed with Diabetes with Chronic Complications in the Traditional Medicare program.

<sup>2</sup>Risk adjustment is used in the US in Medicare Advantage, Medicare Part D, many privatized state Medicaid programs, and the ACA exchanges. It is also widely used throughout European health insurance markets.

Despite wide policy and research interest in “upcoding,” the extent of coding differences across insurers or across the public and privatized arms of programs like Medicare and Medicaid is largely unknown. In most data, upcoding would be observationally equivalent to adverse selection: An insurer might report an enrollee population with higher than average risk scores either because the consumers choosing its plan are in worse health (selection) or because for the same individual, the insurer uses coding practices that result in higher risk scores (upcoding). Because of this central identification difficulty, there has been little empirical work on coding intensity in any US health insurance market. Two important exceptions are [Silverman and Skinner \(2004\)](#) and [Dafny \(2005\)](#), which exploit changes over the 1990s in how Traditional Medicare compensated hospitals on the basis of certain diagnoses, showing that hospital coding patterns responded to track the pattern of reimbursement changes for particular diagnoses. The downside of such difference-in-differences approaches around a group of codes is that while these studies demonstrate clearly that coding behavior endogenously responds to the parameters of the payment system (and for [Dafny, 2005](#) without any change in real service provision), they cannot quantify impacts of coding intensity differences across an entire market. As we show below, differences across insurers or market segments are what drive consumer choice distortions, and in the case of privately provided public insurance like Medicare and Medicaid, impact the bottom line public costs.

To identify coding intensity differences across insurers within a market, we develop a general method for separating upcoding from selection, which is applicable to any selection market. The core insight of our approach is novel, but straightforward: We note that if the same individual would generate a different risk score under two insurers and if we observe a change in market share of the two insurers, then we should also observe changes to the observed *market-level* average of the risk score. Such a pattern could not be rationalized by selection, because selection could affect only the sorting of risk types across health plans within the market, not the overall market-level distribution of reported risk types.<sup>3</sup> Our approach contrasts with past attempts by regulators and researchers to identify upcoding by following individual switchers across plans, which requires assuming that consumers change health plans for reasons unrelated to changes in their health.<sup>4</sup>

---

<sup>3</sup>With this method, the only assumption required to ensure an unbiased estimate of the existence and extent of upcoding is exogeneity of the changes in market share to true underlying population health.

<sup>4</sup>See for example, [Government Accountability Office \(2013\)](#). Our method does not require this identifying assumption. Additionally, results from studies of switchers, unlike results from our method, are often not generalizable to the entire population, again prohibiting estimation of the overall effect of upcoding on the public’s finances.

We exploit large and geographically diverse increases to MA enrollment that began in 2006 in response to the Medicare Modernization Act in order to identify variation in MA penetration that was plausibly uncorrelated with changes in *real* underlying health at the market (county) level. Using the rapid within-county changes in penetration that occurred over our short panel, we find that a 10 percentage point increase in MA penetration leads to a 0.7 percentage point increase in the reported average risk score in a county. Because risk scores have a mean of one, this implies that MA plans generate risk scores for their enrollees that are on average 7% larger than what those same enrollees would have generated under TM.

We show that it is difficult to rationalize our results by the alternative explanation that true population health was changing contemporaneously with these penetration changes within markets. First, the effect is large: a 7% increase in the average risk score is equivalent to 6% of all people becoming paraplegic, 15% of all people contracting HIV, or 58% becoming diabetics.<sup>5</sup> Second, we exploit an institutional feature of Medicare Advantage that causes risk scores to be based on prior year diagnoses. This yields sharp predictions on the timing of effects that are consistent with our findings: Risk scores respond to MA penetration changes with a one year lag, and show no response to penetration changes in the contemporaneous year or any future year. Further, because a portion of the risk scoring algorithm is based on demographic characteristics that are not manipulable by insurers, we verify in a separate falsification test that changes in MA penetration predict only changes to the diagnosis portion of the risk score and not to the demographic portion. Finally, we show that at the county-level MA penetration does not predict other observable time-varying county characteristics that indicate health status but were not plausibly manipulable by insurer coding, including external measures of cancer morbidity, overall mortality, and the age distribution inside and outside of Medicare.

We supplement our main results by investigating the degree to which the level of insurer-provider integration influences a plan's ability to induce intensive coding. Insurers have full knowledge of the risk-scoring algorithms that convert diagnosis codes to payments, but their ability to influence providers to adhere to intensive coding may vary by contract type. For example, HMOs can create risk-based contracts with physician groups that fully pass-through their coding incentives. We show

---

<sup>5</sup>While these effects are large, they are not implausibly large. CMS began deflating Medicare Advantage risk scores by 3.41% in 2010 because of suspected differential coding, while the Government Accountability Office has consistently argued for a larger deflation.

that the coding intensity is significantly higher for managed care MA plans, which are likely to have closer ties to physician groups, than for insurers with less provider integration. Risk scores in managed care MA plans are about 10% larger than they would have been for the same beneficiary in TM, while risk scores in Private Fee-for-Service (PFFS) plans, which did not have physician networks, are only 4-5% larger.<sup>6</sup>

We make four important contributions to the literature on adverse selection and the public finance of healthcare. First, ours is the first paper to model the implications of differential coding patterns across insurers. While there has been substantial research into the statistical aspects of diagnosis-based risk adjustment models in the health services literature, the distortionary implications of coding heterogeneity have received little attention. In the recent applied theoretical work on inefficient selection in insurance markets, risk adjustment has also been largely ignored (see [Einav, Finkelstein and Levin, 2010](#) and [Einav and Finkelstein, 2010](#) for overviews), despite that risk adjustment is the most widely implemented regulatory response to selection.

Our model shows how differences in coding may cause excess public spending and always cause transfers across health plans that distort consumers' choices—in our case between the market segments of Traditional Medicare and Medicare Advantage. A non-obvious result that emerges from our modeling is that for many policy questions regarding the public finance of health insurance and regulatory incidence, it is not necessary to take a stand as to which insurer's coding regime is the “correct” reference coding. In our empirical setting, this means that it doesn't matter whether physicians under TM pay “too little” attention to coding or whether MA insurers pay “too much” attention to coding. This implies that the line between squishy legal coding and illegal “upcoding,” which has been a focal point for some regulators, lacks any economic significance—only relative differences in coding intensity matter for consumer choice and public costs.

Second, we provide a simple and intuitive method for estimating the presence, direction, and extent of coding differences across plans in insurance markets. This method may also be useful for separating selection and program effects in other contexts where, within a geographic market, a fixed population chooses between public and private providers of a service. For example, our method could be used to estimate causal effects of charter schools on student outcomes in a way that is robust to endogenous sorting of students across schools. Our strategy is analogous to the one

---

<sup>6</sup>Employer-sponsored MA plans, which have weaker incentives for upcoding due to a principal-agent problem, also display weaker coding behavior.

used by [Chetty, Friedmand and Rockoff \(2014\)](#), who identify teacher value added via changes to the composition of teaching staff in schools over time, avoiding selection biases by examining results at the school-grade level, rather than at the teacher level.

Third, we provide the first econometric evidence of upcoding in any private insurance market in general, and in MA in particular. While MA enrollees look healthier than TM enrollees because of selection, in reality they are even healthier than they look. While, similar to [Brown et al. \(2014\)](#) who study the costs of cream-skimming in MA, we cannot perform a full welfare analysis of this coding difference, we note that the public spending implications are significant. Medicare is the costliest public health insurance program in the world, and makes up a significant fraction of US government spending. In 2014, the Medicare Advantage program accounted for 30% of total Medicare spending. Absent a coding correction, our estimates imply excess payments of around \$11.4 billion to Medicare Advantage plans annually, or about \$700 per MA enrollee per year. This is about 2 times as large as the [Brown et al. \(2014\)](#) estimate of the increase in excess payments to MA plans due to unpriced selection following the implementation of risk adjustment. This subsidy distorts Medicare beneficiaries' choice of health insurance away from Traditional Medicare, effectively providing a larger voucher for purchasing an MA plan than for purchasing TM. We show that increasing the competitiveness of the MA market, while it would increase consumer surplus, would worsen net efficiency due to this margin of distortion: Competition leads to greater pass-through of this subsidy to enrollees, further distorting consumer choices towards MA.

Finally, our findings contribute to the growing policy literature on the broader welfare impacts of the MA program. Besides the benefits of expanding choice, one popular argument in favor of Medicare Advantage is that it might create important spillover effects on Traditional Medicare. Studies of physician and hospital behavior in response to the growth of managed care suggest the possibility of positive externalities in which the existence of managed care plans lowers costs for all local insurers (see for example, [Baker, 1997](#); [Glied and Zivin, 2002](#); [Glazer and McGuire, 2002](#); [Frank and Zeckhauser, 2000](#)). Most recently, [Baicker, Chernew and Robbins \(2013\)](#) find that the expansion of MA resulted in lower hospital costs in Traditional Medicare. Our findings indicate that these benefits of privatized Medicare do not come without costs. Any positive spillovers should be balanced alongside the additional cost (the deadweight loss of taxation plus welfare losses due to distorted choices) of upcoding in MA.

The outline for the remainder of the paper is as follows. In Section 2, we provide a brief overview of how insurers can influence the diagnoses assigned to their enrollees. In Section 3, we derive a general expression for the implicit subsidy caused by coding differences and present suggestive evidence of relatively intense coding in MA. Section 4 explains our strategy for estimating upcoding in the presence of selection. In Section 5, we discuss our data and empirical setting. Section 6 presents results, and Section 6 discusses several implications of our findings for policy and economic efficiency. Section 7 concludes.

## 2 Risk Adjustment and Endogenous Coding

### 2.1 Risk Adjustment

We begin by briefly describing the functioning of a risk-adjusted payment system in a regulated private insurance market. Plans receive a payment from a regulator for each individual they enroll, which supplements or replaces premiums paid by the enrollee. The risk-adjusted payment  $R$  for enrolling individual  $i$  is equal to the individual's risk score,  $r_i$ , multiplied by some benchmark payment,  $\phi$ , set by the regulator:  $R_i = \phi \cdot r_i$ .<sup>7</sup> The regulator distributes these risk-adjustment payments from a fund, or enforces transfers between plans.<sup>8</sup> The risk score itself is calculated by multiplying a vector of risk adjusters,  $x_i$ , by a vector of risk adjustment coefficients,  $\Lambda$ . Risk adjusted payments are therefore  $R_i = \phi \cdot x_i \Lambda$ .

In health insurance markets, risk adjusters  $x_i$  typically consist of a set of indicators for demographic groups (age-by-sex cells) and a set of indicators for condition categories, which are mapped from the diagnosis codes contained in health insurance claims. In Medicare, these condition categories are referred to as Hierarchical Condition Categories (HCCs). Below, we refer to  $x_i$  as conditions for simplicity. The coefficients  $\Lambda$  capture the incremental impact of the various condition on the insurer's expected costs, as estimated by the regulator in a regression of total cost on conditions in some reference population. They are normalized so that the average risk score is equal to one in the relevant population. The implicit assumption underlying the functioning of risk-adjustment is that conditions  $x_i$  do not vary according to the plan in which a consumer is enrolled. In other words,

---

<sup>7</sup>The benchmark payment can be equal to average *ex-post* realized costs in the full population of enrollees, as in the ACA exchanges, or some statutory amount, as in Medicare Advantage.

<sup>8</sup>The fund can be financed via tax revenues or via fees assessed to health plans by the regulator.

diagnosed medical conditions are properties of individuals, not individual-plan matches. We relax this assumption below and explore the public finance implications.

## 2.2 Endogenous Coding

We define upcoding as differences in coding practices across plans that would lead to two plans generating distinct risk scores for the same individual. Because  $\phi$  and  $\Lambda$  are set by the regulator and fixed across insurers, upcoding can arise only from differences in the reporting of conditions. Formally, we relax the fixed risk scoring assumption by allowing the reported conditions for individual  $i$  to vary by plan  $j$ . The difference in the risk adjusted payment for individual  $i$  between plans  $j$  and  $j'$  is

$$\Delta R_i = \phi \cdot (r_i^j - r_i^{j'}) = \phi \cdot \Lambda(x_i^j - x_i^{j'}). \quad (1)$$

If plan  $j$  codes more intensively than  $j'$  in the sense of reporting more (or more generously reimbursed) diagnoses for an enrollee in the same health state, then  $j$  would receive a larger payment for enrolling the same consumer. In the case of Medicare, if MA insurers generate systematically higher risk scores ( $r_i^{MA}$ ) than does the uncoordinated reporting of diagnoses in the Traditional Medicare system, then Medicare beneficiaries are implicitly provided with a voucher for the purchase of MA that is in excess of the Traditional Medicare voucher by the amount  $\phi \cdot \Lambda(x_i^{MA} - x_i^{TM})$ .

It is important to note that while financial incentives to code are often assumed to drive coding intensity, calculating this difference in voucher size,  $\Delta R_i$ , requires no assumptions about the source of coding differences. This is important because it implies that characterizing the public costs and the consumer choice distortions that arise from differential coding does not require defining some objectively correct level of coding. In our empirical setting below, it doesn't matter whether physicians under Traditional Medicare pay "too little" attention to coding or whether MA insurers pay "too much" attention to coding. As long as the same individual would generate a higher risk score under an MA plan than in TM, the government payout is larger under MA, and consumers will be pushed to the more generously subsidized MA market segment. We discuss the choice distortion implied by this differential subsidy in more detail in Section 6. We also discuss in Section 6 the role of competition: In short, greater competition among MA plans forces greater pass-through of capitation payments to consumers (Cabral, Geruso and Mahoney, 2014), but because the distortion created by upcoding arises between the TM system and the MA system, this competition within MA cannot



correct the potential inefficiency arising from the too-high capitation payments.

## 2.3 Coding in Practice

We discuss briefly what upcoding looks like in practice. The basis for all diagnosis codes is documentation from a provider-patient interaction. During an interaction like an office visit, a physician takes notes, which are passed to the billing/coding staff in the physician's office. Billers use the notes to generate a claim that is sent to the insurer for payment, with diagnosis codes listed on the claim. The insurer pays the claims and over time aggregates all the diagnoses associated with an enrollee to generate a risk score on which a payment from the regulator is based.

Figure 1 outlines the various mechanisms insurers employ to affect diagnosis coding, and in turn, risk scoring.<sup>9</sup> We leave out any mechanisms that involve illegal action on the part of insurers.<sup>10</sup> First, and prior to any patient-provider interaction, insurers can contract with physician groups such that the payment to the group is a fraction of the risk-adjusted payment that the insurer itself receives from the regulator, directly aligning the physicians' incentive toward intensive coding. Insurers may also choose to selectively contract with providers who code more aggressively.

The insurer can influence coding during the medical exam itself by providing tools to the physician that pre-populate his notes with information on prior-year diagnoses for the patient. Since risk adjustment is based solely on the diagnosis of a single prior year, this increases the probability that diagnoses, once added, are retained into the next risk scoring period. Insurers also routinely provide training on how to assign codes to the *physician's* billing staff to ensure coding is consistent with the insurer's financial incentives. Finally, even after claims and codes are submitted to the insurer for a healthcare event, the insurer may automatically or manually review claims, notes and charts, and either request a change to the coding by the physician's billing staff, or directly alter the codes itself.<sup>11</sup>

Beyond these interventions with physicians and their staffs, insurers directly incentivize their enrollees to take actions that result in more intensive coding. Insurers may incentivize or require that enrollees complete annual evaluation and management visits or "risk assessments," which are inexpensive to the insurer, but during which codes can be added that would otherwise have gone

---

<sup>9</sup>Insights in the figure come from investigative reporting by the center for public integrity, statements by CMS, and our own interviews with Medicare Advantage insurers.

<sup>10</sup>X

<sup>11</sup>Insurers use a variety of software tools to scan medical records and determine for each individual the most lucrative combination of codes consistent with the medical record.

undiscovered. Further, if an insurer observes that a potentially high-revenue enrollee has not yet visited a physician in the plan year, the insurer can directly intervene and send a physician or nurse to the enrollee's home. The visit is necessary in order to add the relevant, reimbursable diagnoses for the current plan year but relatively low cost. There is substantial anecdotal evidence for such behavior in Medicare Advantage.<sup>12</sup> And regulators have expressed serious concern that such visits primarily serve to inflate risk scores.<sup>13</sup>

None of these insurer activities take place in Traditional Medicare because providers under the traditional system are paid directly by the government, and in the outpatient setting, these payments are based on procedures, not diagnoses. This difference in incentive structure and between Traditional Medicare and MA naturally suggests that coding will be less intensive under TM, especially with respect to the codes that are relevant for payment in MA. On the other hand, MA insurers are the residual claimants on any healthcare savings they create, so there is a competing incentive to minimize contact between the enrollee and the healthcare system, which could lower the recorded diagnoses.

## 2.4 Suggestive Evidence

One way in which coding may become more intensive when insurers are incentivized to produce higher risk scores is in the reporting of a condition's severity. In Medicare's risk adjustment model, a few common conditions are nested into condition "hierarchies" and then ranked on the basis of severity within the hierarchical group. For example, Table 1 lists the five conditions related to diabetes, ordered by increasing severity. A patient is assigned only the most severe condition for which she has a corresponding diagnosis code within the hierarchy, and more severe conditions generate higher payments. For example, in a county with the average monthly benchmark payment ( $\phi$ ), the incremental increase in the annual MA plan payment generated by moving an enrollee from the least severe diabetes diagnosis (HCC 19) to the most severe diabetes diagnosis (HCC 15) is \$3,263.<sup>1415</sup>

---

<sup>12</sup>Center for Public Integrity 2014.

<sup>13</sup>In a 2014 statement, the Centers for Medicare and Medicaid Services noted that home health visits and risk assessments "are typically conducted by healthcare professionals who are contracted by the vendor and are not part of the plan's contracted provider network, i.e., are not the beneficiaries' primary care providers." CMS also noted there is "little evidence that beneficiaries' primary care providers actually use the information collected in these assessments or that the care subsequently provided to beneficiaries is substantially changed or improved as a result of the assessments."

<sup>14</sup>The average monthly benchmark payment was \$785.85 in 2011. The difference in the risk score between HCC 15 and HCC 19 is 0.346 (= 0.508 - 0.162).

<sup>15</sup>The risk adjustment coefficients for each HCC are shown in Column 3 of Table 1.

Before describing our identification strategy and main dataset, we briefly provide suggestive evidence of the relative coding intensity in Medicare Advantage by examining the distribution of MA and TM codes across these hierarchies. To do so, we present summary statistics from a novel supplementary dataset.<sup>16</sup> We obtained data on the universe of commercial health insurance claims in the state of Massachusetts from 2011 to 2012, which includes claims to Medicare Advantage insurers. Separately, we obtained data from CMS on the universe of Medicare claims for individuals enrolled in Traditional Medicare. For both samples, we used the diagnoses in the claims data to calculate individual-level vectors of conditions ( $\mathbf{x}_i^j$ ) that are reimbursed by Medicare’s risk adjustment scheme.

We use the diabetes hierarchy to provide motivating evidence of upcoding in Medicare Advantage. To create the most comparable sample of TM and MA enrollees, we first isolate all individuals with no diabetes diagnosis in year 1 but at least one diabetes diagnosis in year 2 in both samples. We then calculate the distributions of the year 2 diabetes diagnoses across conditions in the hierarchy separately for TM and MA. The results can be found in the top panel of Table 1. The least severe diabetes condition accounts for most cases in both TM and MA, but more of the mass is distributed toward higher severity cases for MA enrollees.

In the bottom panel of Table 1 we show analogous statistics for the cancer hierarchy as a placebo test. Whereas comparing diabetes diagnoses between MA and TM captures both differences in coding and differences in selection on true underlying health conditions, comparing cancer diagnosis is likely to reveal only health differences, due to the institutional details of how cancer severity is assigned. Cancer diagnoses are grouped according to severity, with higher severity tied to larger payments. But in contrast to diabetes, for which severity is determined presence of potentially manipulatable complicating condition codes, cancer severity is determined simply by the type of cancer.

We condition on Medicare beneficiaries with no cancer diagnosis in year 1 and at least one cancer diagnosis in year 2. Table 1 shows that unlike the distribution of diabetes HCCs, the distributions of cancer HCCs among individuals with no prior cancer diagnosis are very similar in TM and MA. These statistics offer some suggestion that MA patients may be no sicker than TM patients, but may be coded more intensively where coding is squishy, as in the coding of complicating conditions in diabetes.

This example helps to shed light on one way in which coding may differ across insurers or mar-

---

<sup>16</sup>Because the data used to construct the results in Table 1 are different from the data used in our main analysis, we describe the construction of the statistics in more detail in Appendix A.2.

ket segments. Nonetheless, the exercise has important limitations. For one, shifting to higher severity within a condition category may be empirically less important for coding intensity differences than the extensive margin of receiving a diagnosis of any reimbursable condition. Most importantly, an exercise of this type can not rule out that the observed differences in severity in diabetes are entirely due to selection into MA on underlying health. Such tests highlight the inherent problem with identifying coding differences in selection markets. We discuss our solution to this identification problem in the next section.

### 3 Identifying Upcoding in Selection Markets

The central difficulty of identifying upcoding arises from selection on risk scores. At the health plan level, average risk scores can differ across plans competing in the same market either because of coding practice differences for identical patients, or because patients with systematically different health conditions select into different plans. At the individual level, the counterfactual risk score that a person would generate in another plan during the same plan year is unobservable.

In the previous literature, researchers have attempted to identify upcoding by observing how individual risk scores change when consumers switch from one plan to another (CMS, 2010). The critical assumption behind this strategy is that health plan choice is exogenous to health trajectory. Because consumers are more likely to switch health plans in response to a health shock, this assumption is unlikely to hold in practice. An additional shortcoming of this strategy is that it cannot be used to estimate a treatment effect among new enrollees. This shortcoming is non-trivial in our setting given that most consumers choose between TM and MA upon becoming eligible for Medicare at age 65, and their choices exhibit a substantial amount of inertia (Sinaiko, Afendulis and Frank, 2013).

Our solution to the identification problem is to focus on market-level risk. Within a large geographic market, the total population distribution of *actual* health conditions should not change sharply year-to-year. But market-level *reported* risk scores could change sharply if market shares shift rapidly between plans with more and less intense coding practices.

Figure 2 provides the graphical intuition for this idea for the simple case in which selection generates average risk score curves that are linear in market shares. We depict two plans or market segments labeled A and B. These are intended to align respectively with Traditional Medicare and Medicare Advantage. All consumers choose either A or B. Segment B is assumed to be advanta-

geously selected on risk scores, so that the risk score of the marginal enrollee is higher than the average.<sup>17</sup>

In the top panel of Figure 2 we plot the baseline case of no upcoding. The panel shows three curves: average risk in A ( $\bar{r}^A$ ), average risk in B ( $\bar{r}^B$ ), and the average risk of all enrollees in the market ( $\bar{r}$ ). The market share of B, denoted by  $\theta^B$ , is increasing along the horizontal axis. Average risk in B is low at low levels of  $\theta^B$  because the few beneficiaries selecting into B are especially low risk. As long as there is no coding difference between A and B, the market-level risk averaged over enrollees in both plans is constant in  $\theta^B$ . This is because reshuffling of enrollees across plans within a market doesn't affect the market-level distribution of underlying health conditions.

The bottom panel of Figure 2 incorporates differential coding: Segment A is assumed to assign a higher risk score to the same individual compared to segment B. Here, we model the risk score as comprised of a plan-independent individual risk component  $\hat{r}_i$  plus a plan-specific coding factor  $\alpha_j$ :<sup>18</sup>

$$r_i^j = \hat{r}_i + \alpha_j. \quad (2)$$

Differential coding shifts the average risk curve for B by the difference  $\alpha_B - \alpha_A$ . The dashed line in the figure represents the counterfactual average risk that segment B enrollees would have been assigned under segment A coding practices,  $\bar{r}_A^B$ . The bottom panel shows that the unobservable distance between  $\bar{r}_{i \in B}^A$  and  $\bar{r}^A$  identifies the selection effect, while the observable change in market level risk  $\bar{r}$  over the range  $[0, 1]$  of  $\theta^B$  identifies the coding effect.

If and only if there are coding differences between A and B, then the slope of the market-level risk curve with respect to marketshare  $\left( \frac{\partial \bar{r}}{\partial \theta^B} \right)$  will be different from zero. This is true whether or not one of the plans is adversely selected, and regardless of whether the assumption in Eq. 2 holds.<sup>19</sup> Therefore, variation in market share that is exogenous to underlying population health identifies the presence of coding differences.

Under the assumption in Eq. 2 that coding differs by an additive factor, the slope  $\frac{\partial \bar{r}}{\partial \theta}$  is exactly equal to  $\alpha_B - \alpha_A$  for any distribution of risks  $\hat{r}_i$  and for any shape of the selection functions  $\bar{r}^A(\theta)$  and

<sup>17</sup>We ignore uncompensated selection, since our goal here is to distinguish between differences in the (compensated) risk score due to coding and differences in the (compensated) risk score due to selection.

<sup>18</sup>In Appendix ??, we derive this additive, plan-specific factor,  $\alpha_j$ , as being generated from a plan's profit maximization problem.

<sup>19</sup>For completeness, in Appendix Figure ??, we depict the case in which coding differences exists absent any selection on the risk score, and the case in which the adversely selected plan is more intensely coded.

$\bar{r}^B(\theta)$ .<sup>20</sup> In the empirical exercise below, we assume that coding differences take this additive form, since our aggregate market-level data cannot be used to identify heterogeneity in coding effects across individual enrollees. If plans upcode different types of enrollees differently, then the parameters we estimate are local approximations to the mean coding difference.

## 4 Setting and Empirical Framework

We now apply our insight for identifying upcoding in selection markets to estimate the excess public spending and choice distortions that arise due to coding differences between Medicare Advantage (MA) plans and the Traditional Medicare (TM) program. We begin with an overview of the institutional features of payments to private plans in MA. Then we describe our data and discuss our identifying variation and empirical framework in detail.

### 4.1 Medicare Advantage Payments

Individuals who are eligible for Medicare can choose between the TM system administered by the federal government or coverage through a private plan chosen in the MA market. MA plans are attractive to Medicare beneficiaries because compared to the traditional system they offer more comprehensive financial coverage, such as lower deductibles and coinsurance rates, as well as additional benefits, such as dental care and vision care. The tradeoff faced by beneficiaries in choosing an MA plan is that most are managed care plans, which restrict enrollees to a particular network of doctors and may impose referral requirements and other mechanisms to limit access to specialists.

The regulator, the Centers for Medicare and Medicaid Services (CMS), makes monthly capitation payments to MA plans for each beneficiary enrolled. As described in Section 2, the capitation payment ( $\phi_c \cdot r_i^j$ ) is a function of the benchmark rate  $\phi_c$ , which varies across counties  $c$ , and a person-specific adjustment determined by her risk score  $r_i^j$  in the plan, where  $i$  indexes consumers and  $j$  indexes insurers. Historically, county benchmarks have been intended to capture the cost of enrolling the “national average beneficiary” in the Traditional Medicare program in the county, though Congress has made many *ad-hoc* adjustments over time.<sup>21</sup> In 2004, CMS began transitioning from risk adjustment that was based primarily on demographics to risk adjustment based on diagnoses

---

<sup>20</sup>Proof:  $\frac{\partial \bar{r}}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{1}{N} \sum (\hat{r}_i + \alpha_A + \mathbb{1}[B(\theta)](\alpha_B - \alpha_A)) = \alpha_B - \alpha_A$ .

<sup>21</sup>In practice, benchmarks can vary by plan. See Appendix After A.1 for full details.

obtained during inpatient hospital stays and outpatient encounters. By 2007, diagnosis-based risk adjustment was fully phased-in.

Risk scores in Medicare are determined by multiplying risk adjustment coefficients  $\Lambda$  with indicators for groups of diagnoses known as Hierarchical Condition Categories (HCCs). The risk adjustment model also assigns coefficients to age-by-sex cells. CMS sets risk adjustment coefficients nationally using claims data from Traditional Medicare.

## 4.2 Data

Estimating the slope  $\frac{\partial \bar{r}}{\partial \theta^{MA}}$  from Figure 2 requires observing market-level risk scores at varying levels of MA penetration. We obtained county-level averages of MA, TM, and total market risk scores from CMS for 2006 through 2011.<sup>22</sup> MA enrollment is defined as enrollment in any Medicare Advantage plan type, including managed care plans like HMOs and PPOs, Private Fee For Service, employer MA plans, and Special Needs Plans that serve Medicare-Medicaid dual eligibles.<sup>23</sup> Average risk scores within the MA and TM market segment are weighted by the fraction of the year each beneficiary was enrolled in the segment. We define MA penetration as the fraction of all beneficiary-months within a county spent in an MA plan during a given year. For most of our analysis, we collapse all MA plans together, and consider the markets as divided between the MA and TM segments, though we also analyze the partial effects of increased penetration among various subsets of MA plan types.

We supplement these county-level aggregates with administrative data on demographics for the universe of Medicare enrollees from the Medicare Master Beneficiary Summary File (MBSF) for 2006-2011. These data allow us construct county-level averages of the demographic component of risk scores, which we use below in a falsification test.<sup>24</sup>

Table 2 displays summary statistics for the balanced panel of 3,128 counties that make up our analysis sample. The columns compare statistics from the introduction of risk adjustment in 2006 through the last available data year 2011. These statistics are representative of counties, not individuals, since our unit of analysis is the county-year. The table shows that risk scores, which have a mean of approximately 1.0, are lower in MA than overall. While MA enrollees appear healthier than

<sup>22</sup>Similar data are unavailable prior to 2006.

<sup>23</sup>We exclude only enrollees in the Program of All-inclusive Care for the Elderly (PACE) plans.

<sup>24</sup>The demographic components ( $r_i^A$ ) and diagnostic components ( $r_i^{D^x}$ ) of individual risk scores are additively separable, which implies that the county averages of these are also additively separable:  $\bar{r}_i^j = \frac{1}{n} \sum_{i \in I_c} (r_i^A + r_i^{D^x}) = \bar{r}_i^A + \bar{r}_i^{D^x}$ . The  $j$  superscript is suppressed here for simplicity.



TM enrollees, we show below that they are even healthier than they look. The table also shows the dramatic increase in MA penetration over our sample period.

### 4.3 Identifying Variation

#### 4.3.1 MA Penetration Changes

We exploit the large and geographically heterogeneous increases in MA penetration that followed implementation of the Medicare Modernization Act of 2003. The Act introduced Medicare Part D, which was implemented in 2006 and added a valuable new prescription drug benefit to Medicare. Because Part D was available solely through private insurers and because insurers could combine Part D drug benefits and Medicare Advantage insurance under a single contract known as an MA-Part D plan, this drug benefit was highly complementary to enrollment in Medicare Advantage. Additionally, Medicare Advantage plans were able to “buy-down” the Part D premium paid by all Part D enrollees. This led to fast growth in the MA market segment (Gold, 2009). In the top panel of Figure 3, we put this timing in historical context, charting the doubling of Medicare Advantage penetration nationally between 2005 and 2011. The bottom panel of the figure shows that within-county penetration changes were in almost all cases positive, though the size of these changes varied widely. Figure 4 shows that this MA penetration growth was not limited to certain regions or to urban areas. The figure shades each county according to its quantile of penetration changes.

Our identification strategy relies on year-to-year variation in penetration within geographic markets to trace out the slope  $\frac{\partial \bar{r}}{\partial \theta^{MA}}$ . The identifying assumption is that these changes in MA enrollment are not correlated with changes in actual underlying population health. In particular, in the county fixed effects models we estimate below, this implies that year-to-year growth in MA enrollment in the county did not track year-to-year variation in the actual population-level health of the county. The assumption is plausible, given that changes in county population health, reflected in the incidence of chronic conditions such as diabetes and cancer, is unlikely to change sharply year-to-year, while changes in *reported* risk that are due to coding practices can change instantaneously when a large fraction of the Medicare population moves to the MA segment.<sup>25</sup> Further, we can test the assump-

---

<sup>25</sup>We offer the following additional arguments for the plausibility of this identifying assumption. On the supply side, the assumption implies that insurers don’t selectively enter counties or alter plan benefits based on year-to-year changes in the average health of the county. This seems sensible, given that the dramatic penetration growth over our period appears to be driven by regulatory changes to Medicare embodied in the Medicare Modernization Act of 2003. We would spuriously estimate upcoding effects in MA only if insurers expanded market share by lowering prices or increasing benefits in places



tion of no correlated underlying health trends with respect to a variety of independently observable demographic, morbidity, and mortality outcomes at the county level.

In our preferred specification, we use all of the within-market, over-time variation in MA penetration to identify the parameter  $\frac{\partial \bar{r}}{\partial \theta^{MA}}$ . However, we also estimate versions of the regression that isolate different subsets of the year-to-year variation in MA enrollment. In one version, we control for changes in MA penetration that were due to any type of plan other than MA-Part D plans, which isolates the MA penetration growth directly attributable to growth in the MA-Part D market. In a second version, we control for changes in MA penetration that were due to only MA-Part D plans, isolating MA penetration growth that was orthogonal to MA-Part D growth.

#### 4.3.2 Timing

We also exploit an institutional feature of how risk scores are calculated in Medicare Advantage to more narrowly isolate the identifying variation that arises from post-Medicare Modernization Act increases in enrollment. Because risk scores are calculated based on the prior year's diagnoses, up-coding should only be apparent with a lag relative to penetration changes.

We illustrate the timing in Figure 5. The individual's risk score that is used for payment throughout the calendar year  $t+1$  is based on diagnoses from calendar year  $t$ . This implies, for example, that if an individual moves to MA from TM, the risk score for her entire first year in MA will be based on diagnoses she received while in TM in the prior plan year. Only after the first year of MA enrollment will the risk score of the switcher include diagnoses she received while enrolled with her MA insurer. Therefore, in the first year following a net change in MA enrollment due to switching, the overall market level risk should remain constant.

The timing is slightly more complex for new Medicare eligibles choosing to enroll in MA. During the first year of an individual's Medicare eligibility—usually at age 65—her risk score is based only on demographics, since CMS does not have the information on prior diagnoses necessary to construct the full diagnosis-based risk score. This implies that beneficiaries won't have the full calendar year of diagnosis codes required by CMS to create a risk score until their second full year (third calendar year) of MA enrollment, as illustrated in Figure 5. Therefore, changes in MA penetration due to

---

where the population was simultaneously becoming sicker or older. In terms of consumer choice, our assumption implies that individuals' demand for MA does not increase as the average health in the county declines. This seems plausible, given that the literature suggests that if MA enrollees differ at all from TM enrollees, MA enrollees are healthier, not sicker (Brown et al. (2014)).

newly-eligible aged enrollees should appear with a 2-year lag. Why exploit these timing features below.

#### 4.4 Econometric Framework

The slope of market-level average risk with respect to MA penetration identities coding intensity in MA relative to TM. To control for any unobserved local factors that could simultaneously affect population health and MA enrollment, such as physician practice styles, medical infrastructure, or consumer health behaviors, we exploit the panel structure of our data and estimate fixed effects models of the form:

$$\bar{r}_{sct} = \gamma_c + \gamma_t + \left( \sum_{\tau \in T} \beta_{\tau} \cdot \theta_{sct}^{MA} \right) + f(X_{sct}) + \epsilon_{sct}, \quad (3)$$

where  $\bar{r}_{sct}$  is the average market-level risk in county  $c$  of state  $s$  at time  $t$  and  $\theta^{MA}$  denotes MA penetration, which ranges from zero to one. County and year fixed effects are captured by  $\gamma_c$  and  $\gamma_t$ , and  $X_{sct}$  is a vector of time-varying county characteristics described in more detail below. The subscript  $\tau$  indicates timing relative to the reported risk score. Coefficients  $\beta_{\tau}$  multiply contemporaneous MA penetration ( $\tau = t$ ), leads of MA penetration ( $\tau > t$ ), and lags of MA penetration ( $\tau < t$ ).

The coefficients of interest are  $\beta_{t-1}$  and  $\beta_{t-2}$  because of the institutional feature described above in which risk scores are calculated based on the prior full year's medical history, so that upcoding could plausibly affect risk scores only after the first year of MA enrollment for prior TM enrollees and after the second year of MA enrollment for newly-eligible beneficiaries. Because of the short panel nature of the data—our data starts in 2006, with usable data beginning in 2007—in our main specification, we estimate  $\beta$  for only a single lag. Later, we report alternative specifications that include a second lag, though these necessarily decrease the sample size, limiting statistical precision. A positive coefficient on lagged penetration indicates more intensive coding in MA relative to TM. Under our additive coding intensity assumption in Eq. 2,  $\beta$  is exactly equal to  $\alpha_{MA} - \alpha_{TM}$ , which is the difference between the risk scores that the same individual would generate under the two systems.

We include the placebo regressor  $\theta_{sc, \tau=t}^{MA}$  in all specifications. Because upcoding can plausibly affect market level risk only with a lag, the contemporaneous effect of penetration changes on market

level risk reflected in  $\beta_t$  should be zero. The coefficient on the placebo reveals any source of contemporaneous correlation between MA penetration and unobservable determinants of county risk that could contaminate our results. Similar placebo tests can be performed for leads of penetration ( $\theta_{sc\tau>t}^{MA}$ ), again subject to the caveat of reducing the panel length.

Besides these placebo tests, we perform a series of falsification tests, described below, to show that at the county level, MA penetration does not predict other time-varying county characteristics. Most importantly, we test whether MA penetration changes are correlated only with the portion of the risk scores derived from diagnoses. Risk scores are partly determined by demographic characteristics, which are not plausibly manipulated by insurer behavior.

## 5 Results

We begin by presenting the results that include all MA plan types. After reporting on a series of falsification and placebo tests in support of our identifying assumption, we examine how upcoding effects vary according to the level of vertical integration between the insurer and physician.

### 5.1 Main Results

Table 3 reports our main specifications. The coefficient of interest is on lagged MA penetration. In column 1 we present estimates of the baseline model controlling for only county and year fixed effects. The coefficient indicates that the total average risk score a county increases by 0.07—approximately one standard deviation—as lagged MA penetration increases from 0% to 100%. Because risk scores are scaled to have a mean close to one in the population, this also implies that an individual’s risk score in MA is about 7% higher than it would have been under Traditional Medicare (TM). In column 2 we add linear state time trends, and in column 3 we add time-varying controls for county demographics.<sup>26</sup> Across specifications, the coefficient on lagged MA penetration is stable.

To put the size of these coding effects in context, an increase in market-level risk of 0.07 would imply that starting from a perfectly healthy county population, 6% of all people became paraplegic, or 15% all people contracted HIV, or 58% became diabetics. If, contrary to our identifying assumption, these estimates were simply capturing spurious correlation between actual changes in underlying

<sup>26</sup>These controls consist of 19 variables capturing the fraction of population in the county-year in 5 year age bins from 0 to 85.

health conditions in the local market and changes to MA penetration, it would require substantial shocks to the market-level averages of health conditions. Further, it would require that enrollment decisions, which are made in the fall of the year prior to enrollment ( $t - 2$  in the notation Eq. 3), predict these health shocks, conditional on enrollment in adjacent years.

While these effects are large, they are not inconsistent with widely held beliefs about coding in Medicare Advantage. Since 2010, the Medicare regulator, the Centers for Medicare and Medicaid Services, has applied a 3.41% deflation factor when determining payments to private plans in the Medicare Advantage program, under the assumption that private plans code the same patients more intensively. The Government Accountability Office has expressed concerns that coding differences between MA and TM are likely much higher, in the range of 5% to 7%. Neither agency—nor any other study—has been able to provide econometrically identified estimates of this coding difference.

As a robustness check, we isolate the component of our identifying variation that arises explicitly from the expansion of MA-Part D plans, which combined Medicare Advantage with the new prescription drug benefit introduced in 2006. To do so, in columns 4 through 6 of Table 3, we control for changes in MA penetration in any plan type other than MA-Part D. These regressions are identified solely from changes in MA-Part D enrollment, which expanded rapidly in the few years following Part-D's introduction. In columns 7 through 9 of Table 3, we isolate the complement of this variation by controlling for changes in MA-Part D penetration separately from overall MA penetration. All results are closely consistent with our preferred specification in column 3, which uses all within-county across-time variation in penetration.

## 5.2 Placebo Tests

The coefficient estimates for contemporaneous MA penetration in Table 3, which are close to zero and insignificant across all specifications, support our placebo test. These coefficients imply that the health of the population was not drifting in a way that was spuriously correlated with changes in penetration. In principle, we could extend the placebo test of our main regressions by examining leads in addition to the contemporaneous effect. In practice, we are somewhat limited by our short panel, which becomes shorter as more leads or lags are included in the regression. For example, due to the length of time the program has existed, our data extend back only to 2006. The most recent data year available is 2011. Therefore, including two leads and one lag of penetration restricts our panel

to just 2007 to 2009. Nonetheless, in columns 1 through 3 of Table 4, we repeat the main analysis with additional leads, under the intuition that significant coefficients on contemporaneous effects or leads would provide evidence of confounding trends.

Column headers in Table 4 describe the panel years, which necessarily change across columns. Standard errors increase due to the smaller sample sizes, but the patterns on the placebo variables ( $\theta_t^{MA}$ ,  $\theta_{t+1}^{MA}$ , and  $\theta_{t+2}^{MA}$ ) in columns 1 through 3 show no evidence that contemporaneous or future values of penetration are correlated with market-level changes in time  $t$  risk scores. Because true population characteristics tend to change gradually rather than discretely, the precisely timed response with a lag of at least one-year is more consistent with a mechanical coding effect than an impulse change in true population health.

As discussed in the context of Figure 5, coding differences in the Medicare Advantage context should be observable only with one or two year lag: switchers from TM to MA carry forward their old risk scores for one plan-year, and newly-eligible consumers aging into Medicare and choosing MA won't have risk scores based on diagnoses assigned in MA until after two plan years.<sup>27</sup> Column 4, which includes a second lag provides evidence consistent with this. Each coefficient in the table represents an independent effect, so that point estimates on the first and second lag of penetration in column 4 indicate a cumulative upcoding effect of 9.2% after two years. Unfortunately, in the short panel, we are restricted from looking at effects with longer lags or leads with any precision. Nonetheless, we report on an extended set of leads and lags in Appendix Table A1.

### 5.3 Falsification Tests

In Tables 5 through 7 we conduct a series of falsification tests intended to uncover any correlation between changes in MA penetration and changes in other time-varying county characteristics. In particular, we focus on county demographics, mortality, and morbidity, since correlation between these characteristics and MA penetration could undermine the identifying assumption.

Table 5 replicates the specifications in columns 1 through 3 of Table 3, but uses the demographic portion risk scores as the dependent variable. The demographic portion of the risk score is based only on age and gender, and unlike diagnoses is not manipulable by the insurer except by outright

---

<sup>27</sup> Additionally, some of the insurer strategies for coding, such as prepopulating physician notes with past diagnoses and making home health visits to enrollees who had been previously coded with generously reimbursed conditions, would suggest that upcoding effects ratchet up the longer an individual is enrolled in MA. Even for switchers from TM, this could result in positive coefficients for more than a single lag of MA penetration.

fraud. The coefficients, which are near zero and insignificant in all specifications, show no impact of lagged penetration, consistent with the mechanism we describe in which enrollees are assigned more or more severe medical conditions.<sup>28</sup>

In Table 6, we next test whether changes in MA penetration are correlated with measures of mortality and morbidity. Columns 1 through 3 show the relationship between changes in a county's mortality rate and changes in MA penetration. For morbidity, it is important that data are not potentially influenced by insurer coding. We use cancer incidence data from the Surveillance, Epidemiology, and End Results (SEER) Program by the National Cancer Institute. Columns 4 through 6 shows the relationship between cancer incidence in a county-year and MA penetration. Cancer data is limited to the subset of counties monitored by SEER, which accounted for 27% of the US population in 2011, and 25% of the population over 65. Coefficients on both contemporaneous and lagged MA penetration are consistently close to zero and statistically insignificant.

Finally, we test whether changes in MA penetration are correlated with changes in the county's Medicare enrollee age distribution. In Table 7, the dependent variables measure the fraction of a county's Medicare population within each specified age range. The estimates show no consistent evidence of a systematic relationship between MA penetration and the Medicare enrollee age distribution. In sum, each falsification tests supports our identifying assumption of no correlation between MA penetration and actual underlying population health or demographics, conditional on our controls.

## 5.4 Heterogeneity by Insurer-Provider Integration

We have so far treated all MA plans symmetrically, but the power insurers can exert over coding is likely to vary depending on the level of vertical integration between the insurer and the physician groups with whom the insurer contracts. Plans that follow the standard HMO, POS, or PPO models of managed care are likely to be strongly vertically integrated, including paying their physicians partly or wholly as a function of the risk score that the physician's diagnoses generate. Private fee for service (PFFS) plans are fundamentally different. During most of our sample period PFFS plans did not have networks of providers. Instead they reimbursed Medicare providers on the basis of

---

<sup>28</sup> A alternative valid interpretation of the results in Table 5 is that conditional on county fixed effects, MA plans were not differentially entering counties in which the population structure was shifting to older ages, which are more generously reimbursed in the risk adjustment formula.

procedure codes (not diagnoses) at standard Medicare rates. Thus PFFS plans had access to only a subset of the tools available to managed care plans for influencing the recording of diagnoses within the physician’s practice. In particular, PFFS insurers could not arrange a contract with providers that directly rewarded intensive coding, nor would PFFS insurers be likely to train a physician’s billing staff on coding.

In Table 8 we separate out the effect of PFFS plans and “Risk” plans, which nest HMOs, PPOs, and POSs.<sup>29</sup> We also separately control for employer-sponsored MA plans. Employer plans are less strongly incentivized to code intensively as in these, financial insurance is provided by the employer with claims and diagnosis-reporting administered by third-parties. This adds a layer of principal-agent problems into the coding incentive structure.<sup>30</sup>

Consistent with expectations, we find the strongest upcoding effects among managed care plans, with smaller effects for PFFS and employer plans. Table 8 shows that risk scores in managed care plans are around 10% higher than they would have been in TM. Risk scores in PFFS and employer plans, however, are only around 4-5% higher.

## 6 Discussion

### 6.1 Public Finance Impacts

The implicit subsidy to Medicare Advantage (MA) due to differences in coding intensity is equal to the upcoding factor multiplied by the county benchmark rate ( $\phi(\alpha_{MA} - \alpha_{TM})$ ). In 2014, the average annual value of  $\phi$  was about \$10,000. Given our estimate of  $\alpha_{MA} - \alpha_{TM} = 0.07$  for MA overall, this implies a subsidy of about \$700 per MA enrollee in 2014, or a total potential excess subsidy of about \$11.4 billion, absent any coding inflation factor.<sup>31</sup> In 2010, CMS put in place a 3.41% inflation factor. Our results suggest this is both too small and fails to account for large coding differences across insurance contract types. While PFFS plans and employer MA plans differ in coding intensity by 4-5% relative to Traditional TM, HMOs and PPOs, which comprise the largest category of MA plans, inflate risk scores by 10%. This translates to an incremental implicit subsidy of about \$1,000 per MA

<sup>29</sup>Our data do not allow us to disaggregate this category further.

<sup>30</sup>A typical MA insurer only needs to exert its influence on providers to increase coding intensity. However, an employer has to influence the third party administrator to influence providers, likely a more difficult task.

<sup>31</sup>Based on September 2014 enrollment of 16,347,808 beneficiaries in MA, reported in Monthly SCC Enrollment Files provided by CMS.

enrollee annually to HMOs and PPOs.

These costs of differential coding are substantial. [Brown et al. \(2014\)](#) study the uncompensated advantageous selection into Medicare Advantage caused by the implementation of risk adjustment in 2006. They show that the introduction of risk adjustment led to \$317 per enrollee in annual overpayments to MA, relative to the cost of insuring beneficiaries under TM, because MA plans attracted relatively low cost enrollees conditional on their risk scores. Our results imply that the [Brown et al. \(2014\)](#) estimate, while large and important in itself, dramatically understates the problem of implicit overpayments to MA plans arising from risk adjustment because that study doesn't consider heterogeneous coding.

Recent changes to Medicare that are intended to improve quality of care in the Traditional system are likely to have the unintended consequence of encouraging higher coding intensity by Traditional Medicare providers. Interestingly, this might have the benefit of reducing excess payments to MA plans. Newly established Accountable Care Organizations (ACOs) under Traditional Medicare are intended to incentivize cost savings by making providers the residual claimants on healthcare savings of their patients. In the Medicare Shared Savings Program, the most popular ACO program, regulators compensate TM physicians and hospitals when patients under their care use relatively fewer resources relative to their *risk adjusted* health state, potentially incentivizing more intense coding in order to maximize payments under this metric.<sup>32</sup>

Higher coding intensity in Traditional Medicare relative to the status quo could actually reduce the budgetary impacts of intensive coding in MA. Because the relevant diagnoses for ACO reimbursements overlap with those in the MA risk adjustment scheme, coding incentives would be more closely aligned between MA and TM. From Eq. 1, more intense coding under TM would reduce the subsidy wedge between TM and MA by the amount  $\phi(\alpha_{TM}^{new} - \alpha_{TM}^{old})$ .<sup>33</sup> Such a possibility illustrates our claim that most public finance consequences of “upcoding” are completely determined by relative differences in coding intensity. The notion of an objectively correct level of coding has little economic content.

---

<sup>32</sup>It is possible in principle that ACO program could cause coding intensity in TM to surpass the level of intensity in MA due to the removal of the principal-agent complexity inherent in relationship between the MA insurers who are incentivized to upcode and the providers who ultimately do most of the coding.

<sup>33</sup>In order for this change in TM coding intensity to impact overpayments to MA plans due to MA upcoding, CMS will have to recalibrate the MA risk adjustment coefficients using TM data from the post-ACO period.



## 6.2 Welfare Consequences

While we document the impact on public spending, it is difficult to take a stance on the welfare consequences of heterogeneous coding, largely because it is difficult to establish whether increased coding intensity has any direct welfare impacts. For instance, if insurers exert their influence on providers to document their patients' conditions, information-sharing across healthcare providers could be facilitated in a way that impacts patient health.

While this paper cannot systematically address the possibility that intensive coding itself represents a good, we expect such impacts to be relatively small for several reasons. For one, insurers primarily intervene with physicians' coding and billing staff rather than the physicians themselves. Insurers influence the coding of diagnoses in *insurance claims*, to which the physician, who gleans information instead from patient charts, may have no exposure. Second, coding aimed at maximizing payment and coding aimed at improving quality and continuity of care are likely to be different. One large MA insurer explained to us that the type of documentation physicians desire for clinical reasons is often at odds with what insurers desire for risk-scoring.<sup>34</sup> Finally, the regulatory agencies that oversee risk adjustment programs express serious doubt about the clinical value of insurer activities surrounding coding, some of which lack any connection to the actual provision of healthcare. For example, in a 2014 statement, the Centers for Medicare and Medicaid Services explained its view that home health visits and risk assessments "are typically conducted by healthcare professionals who are contracted by the vendor and are not part of the plan's contracted provider network, i.e., are not the beneficiaries' primary care providers."

If intensive coding doesn't meaningfully affect consumer wellbeing by influencing real care provision, then the primary welfare consequences of coding intensity differences are the distortions these introduce into health plan choice. In our empirical setting, seniors are presented with an implicit voucher to purchase an MA plan that is significantly larger than the value of their implicit TM voucher. Similarly, within the MA market segment choices are distorted towards the plans and plan types with higher coding intensity, namely non-employer managed care plans.

A full welfare analysis of the TM/MA choice margin is beyond the scope of this paper, though distorting consumer choices towards MA via this subsidy could be efficient if (i) frictions caused

---

<sup>34</sup>Source: Our interview with an anonymous large managed care MA insurer on April 4, 2014.

lower than optimal enrollment in MA based on private benefits,<sup>35</sup> or (ii) MA enrollment has important external benefits, such as the cost control spillovers documented by [Baicker, Chernew and Robbins \(2013\)](#). In any case, understanding the size of the implicit coding subsidy that we calculate here would be key to any full welfare analysis.

Finally, it is important to recognize that stronger competition within the MA market would have no impact on the excess public costs of upcoding. Competition can't impact the bottom line of public spending because neither the risk adjustment coefficients ( $\Lambda$ ) nor the regulator's benchmark ( $\phi$ ), which together with coding differences determine the size of the overpayment, vary with competition. Competition within the MA market segment determines the allocation of overpayments to MA between providers and consumers, but has no bearing on the size of those overpayments.

In terms of consumer choice, stronger MA competition may paradoxically exacerbate the sorting distortion. If competition is stronger, then insurers will pass-through the coding subsidy to consumers in the form of lower premiums or additional benefits ([Cabral, Geruso and Mahoney, 2014](#)).<sup>36</sup> While this increases consumer surplus at the cost of producer surplus, it may also decrease net efficiency by distorting choices. On the other hand, if competition is imperfect, insurers will pass-through a smaller fraction of the overpayment, resulting in a smaller distortion to seniors' choice between TM and MA. Thus, competition results in a tradeoff between the incidence of the subsidy and net efficiency: If competition is strong, the implicit coding subsidy will result in additional consumer surplus but distorted choices; if competition is weak, the subsidy will result in additional producer surplus but a smaller choice distortion.

In other regulated markets where private insurers don't compete against a public option, such as the ACA Exchanges, upcoding nonetheless remains a significant problem. In these markets, risk adjustment has no first order impact on public budgets, because a regulator simply enforces transfers from plans with lower average risk scores to plans with higher average risk scores.<sup>37</sup> In these settings, plans are incentivized to code intensively to maximize the profit, trading off the incremental subsidy

---

<sup>35</sup>For example, inertia to remain in Traditional Medicare.

<sup>36</sup>Because the TM subsidy is very large and MA plans cannot offer negative premiums, MA plans may be unable to charge a price low enough to induce efficient sorting. However, MA plans are allowed to (and often do in practice) "buy down" seniors' Part B and Part D premiums. This is economically equivalent to offering a negative premium. See the appendix for additional information about the MA payment system

<sup>37</sup>In markets such as the Exchanges where the government pays subsidies based on the premiums set by insurers, there will still be public finance consequences from upcoding. Investment by health plans in coding intensity could conceivably result in higher premiums and, because government subsidies are based on these premiums, higher subsidies and additional government spending.

to intensive coding against its cost. If coding intensity has no direct impact on consumer welfare, then investment in coding practices implies a deadweight loss. Perhaps more importantly, if there is heterogeneity across plans in the cost of increasing coding intensity, then upcoding-induced choice distortions will persist even in fully competitive, fully private insurance markets. Our result that MA plans with higher levels of insurer-provider integration display higher coding intensity suggests that the choices of Exchange enrollees are likely to be inefficiently distorted toward these more integrated plans.

## 7 Conclusion

In this paper we developed a method for identifying upcoding in selection markets, and then used it to evaluate coding differences between Traditional Medicare and Medicare Advantage, the largest risk-adjusted health insurance market in the US. Our findings indicate large differences in coding intensity between Medicare's public and private option, with significant implications for overpayments to private insurers. We also find strong evidence that coding intensity is increasing in a plan's level of insurer-provider integration.

Risk adjustment addresses an important problem of asymmetric information in insurance markets. Therefore, in the second-best world in which adverse selection is an inherent feature of competitive insurance markets, the optimal payment mechanism may include some kind of risk adjustment despite the costs and distortions of manipulable coding that we document. Nonetheless, our study offers some insight into improvements in risk adjustment mechanism design: From the perspective of this paper, the risk adjustment literature focusing on the predictive content of risk scores is pursuing the wrong objective function. [Glazer and McGuire \(2000\)](#) show that to induce efficient sorting, risk adjustment must focus on insurer incentives, rather than maximizing prediction of expected costs. Applied to our findings, this suggests that the (second best) optimal payment policy may include risk adjustment, but with coefficients on diagnoses that account for both predictiveness of costs and for susceptibility to coding intensity differences. In principle, with information on the upcoding susceptibility of various conditions, it would be possible to estimate optimal payment coefficients by minimizing a loss function that included coding distortions. In practice with limited information, fewer and coarser diagnosis groups might be the feasible alternative that is preferable to the current risk adjustment systems. It is an issue of significant practical importance, given the large and grow-

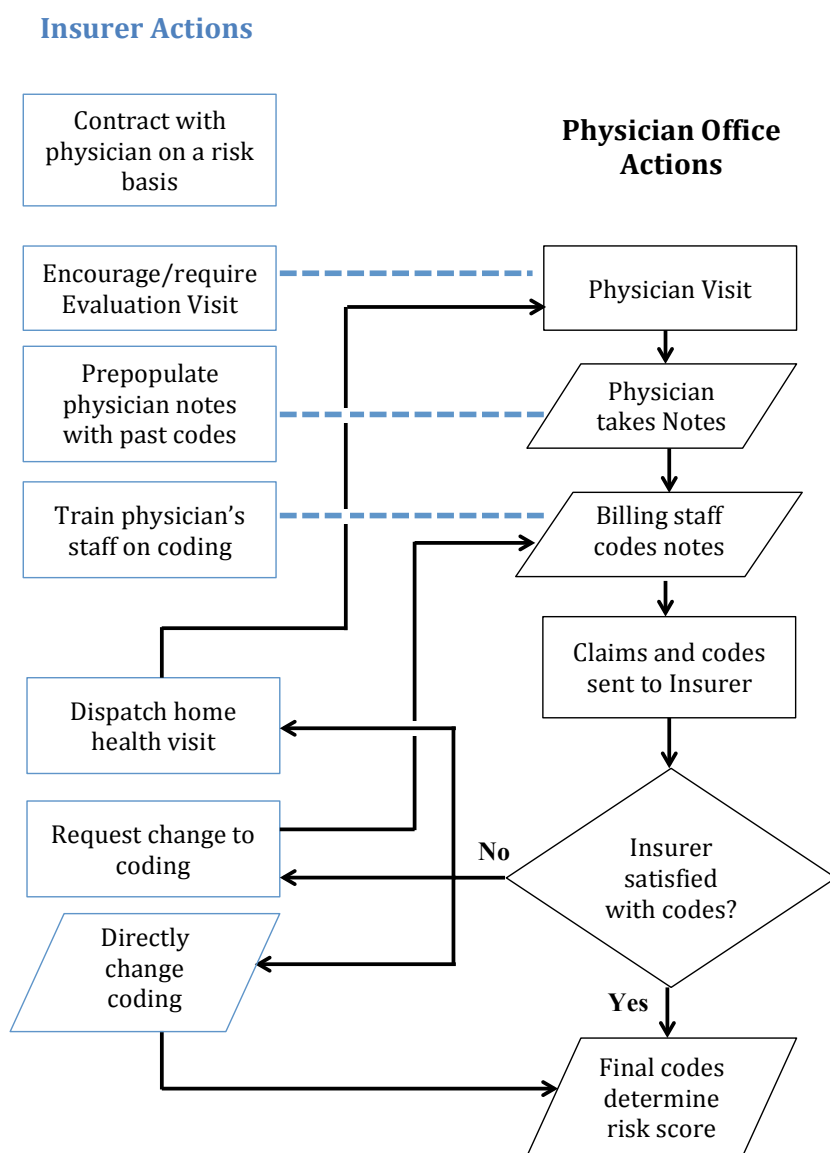
ing role of risk adjustment in regulated insurance markets for Medicare, Medicaid, and the ACA Exchanges.

## References

- Akerlof, George A.** 1971. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics*, 84(3): 488–500.
- Baicker, Katherine, Michael E. Chernew, and Jacob A. Robbins.** 2013. "The spillover effects of Medicare managed care: Medicare Advantage and hospital utilization." *Journal of Health Economics*, 32(6): 1289 – 1300.
- Baker, Laurence.** 1997. "HMOs and Fee-For-Service Health Care Expenditures: Evidence from Medicare." *Journal of Health Economics*, 16: 453–482.
- Brown, Jason, Mark Duggan, Ilyana Kuziemko, and William Woolston.** 2014. "How does risk selection respond to risk adjustment? Evidence from the Medicare Advantage program." *American Economic Review*, 104(10): 3335–3364.
- Cabral, Marika, Michael Geruso, and Neale Mahoney.** 2014. "Does Privatized Health Insurance Benefit Patients or Producers? Evidence from Medicare Advantage." *NBER Working Paper*.
- Chetty, Raj, John N. Friedmann, and Jonah E. Rockoff.** 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.
- CMS.** 2010. "Announcement of Calendar Year (CY) 2010 Medicare Advantage Capitation Rates and Medicare Advantage and Part D Payment Policies." Centers for Medicare and Medicaid Services.
- Dafny, Leemore S.** 2005. "How Do Hospitals Respond to Price Changes?" *American Economic Review*, 1525–1547.
- Einav, L, and A Finkelstein.** 2010. "Selection in Insurance Markets: Theory and Empirics in Pictures." *Journal of Economic Perspectives*, 25(1): 115–138.
- Einav, Liran, Amy Finkelstein, and Jonathan Levin.** 2010. "Beyond Testing: Empirical Models of Insurance Markets." *Annual Review of Economics*, 2: 311–36.
- Einav, Liran, Amy Finkelstein, and Mark R. Cullen.** 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices." *The Quarterly Journal of Economics*, 125(3): 877–921.
- Frank, Richard G, and Zeckhauser.** 2000. "Custom-made versus ready-to-wear treatments: Behavioral propensities in physicians' choices." *Journal of Health Economics*, 26(6): 1101–1127.
- Glazer, Jacob, and Thomas G McGuire.** 2000. "Optimal risk adjustment in markets with adverse selection: an application to managed care." *The American Economic Review*, 90(4): 1055–1071.
- Glazer, Jacob, and Thomas G McGuire.** 2002. "Setting health plan premiums to ensure efficient quality in health care: minimum variance optimal risk adjustment." *Journal of Public Economics*, 84(2): 153–173.
- Glied, Sherry, and Joshua Graff Zivin.** 2002. "How do doctors behave when some (but not all) of their patients are in managed care?" *Journal of Health Economics*, 21(2): 337–353.
- Gold, Marsha.** 2009. "Medicare's Private Plans: A Report Card On Medicare Advantage." *Health Affairs*, 28(1): w41–w54.

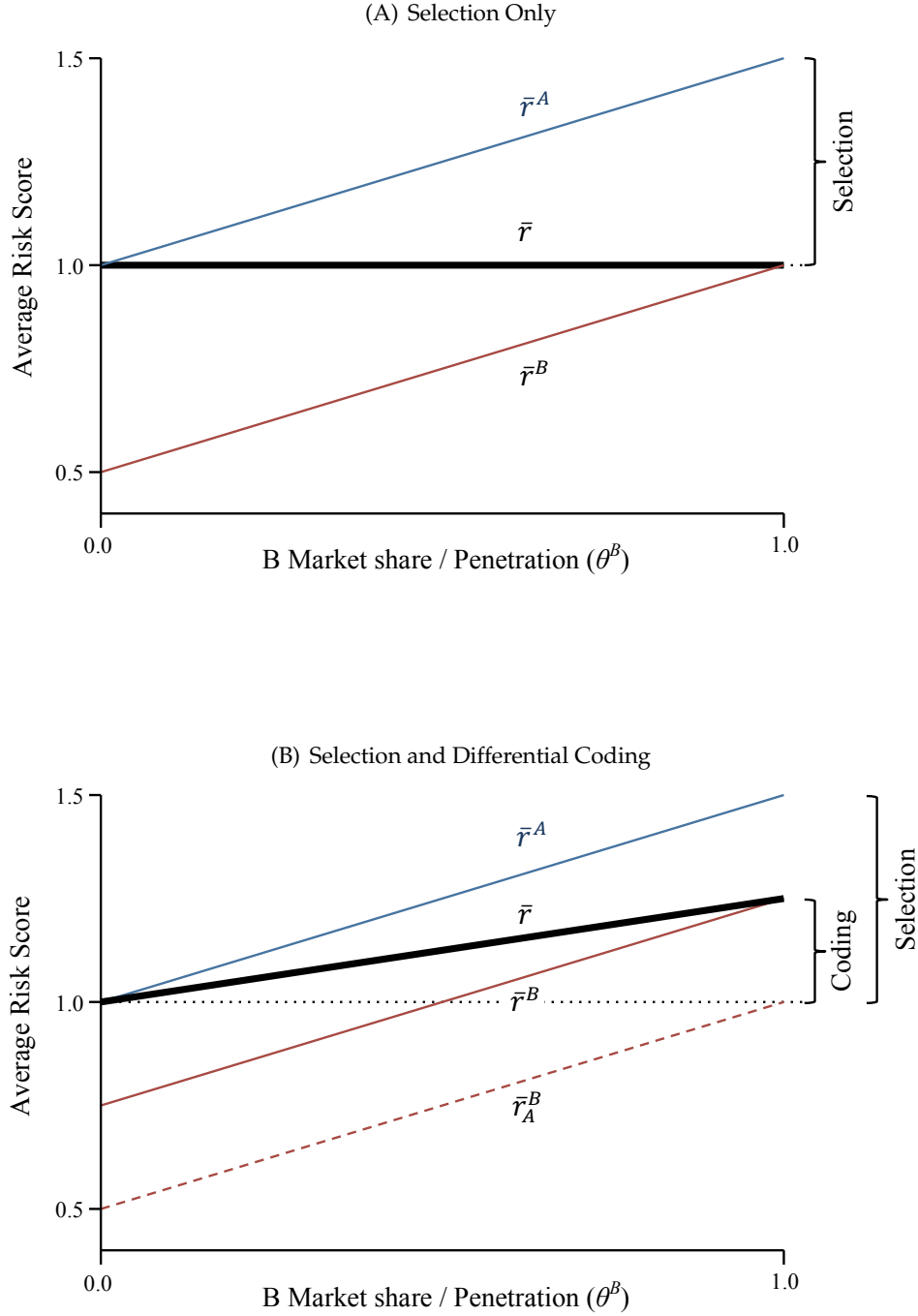
- Government Accountability Office.** 2013. "Substantial Excess Payments Underscore Need for CMS to Improve Accuracy of Risk Score Adjustments." Accessed on Oct 8, 2013 at <http://www.gao.gov/assets/660/651712.pdf>.
- Rothschild, Michael, and Joseph Stiglitz.** 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *The Quarterly Journal of Economics*, 90(4): 629–649.
- Silverman, Elaine, and Jonathan Skinner.** 2004. "Medicare upcoding and hospital ownership." *Journal of health economics*, 23(2): 369–389.
- Sinaiko, Anna D., Christopher C. Afendulis, and Richard G. Frank.** 2013. "Enrollment in Medicare Advantage Plans in Miami-Dade County: Evidence of Status Quo Bias?"
- Song, Yunjie, Jonathan Skinner, Julie Bynum, Jason Sutherland, John E Wennberg, and Elliott S Fisher.** 2010. "Regional variations in diagnostic practices." *New England Journal of Medicine*, 363(1): 45–53.

**Figure 1: How Risk Scores are Influenced by Insurers**



**Note:** The flowchart illustrates how diagnosis codes originate and how insurers can influence the process that generates them. Insurer actions are towards the left of the figure in blue boxes. Provider actions, including those of the provider's billing and coding staff are towards the right in black boxes. Actions that immediately result in code data generation are represented by rhombuses.

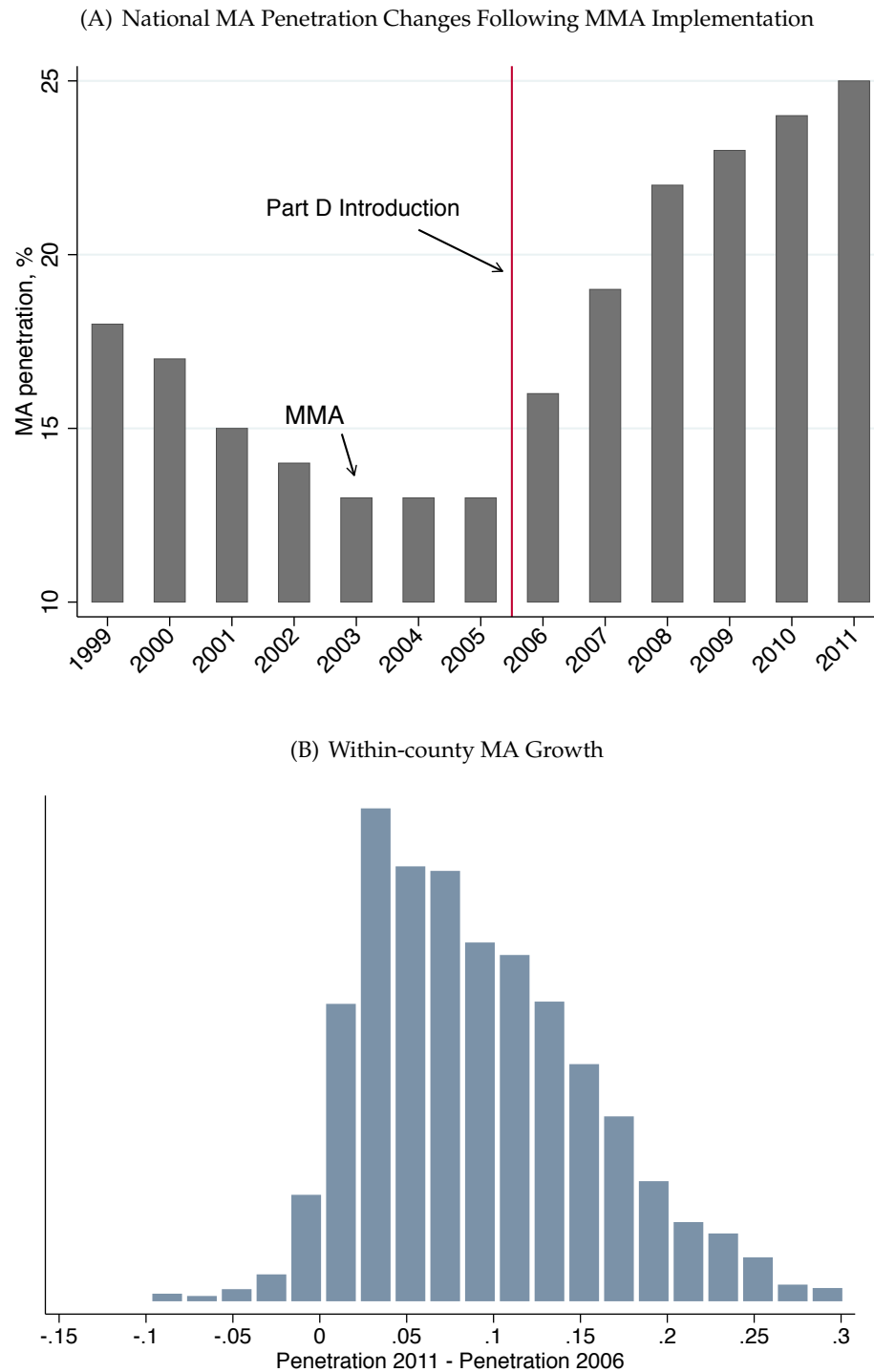
**Figure 2: Identifying Coding Differences in Selection Markets**



**Note:** This figure demonstrates how to separate coding differences from selection when true underlying risk is unobservable. The horizontal axis measures the market share of segment  $B$ ,  $\theta^B$ . The vertical axis measures the average risk score: Average risk in  $A$  is  $\bar{r}^A$ , average risk in  $B$  is  $\bar{r}^B$ , and the average risk of all enrollees in the market is  $\bar{r}$ . The dashed line in the figure represents the counterfactual average risk that segment  $B$  enrollees would have been assigned under segment  $A$  coding practices,  $\bar{r}_A^B$ . All consumers choose either  $A$  or plan  $B$ . Segment  $B$ , which models Medicare Advantage, is assumed to be advantageously selected in both panels, and additionally is assumed to have higher coding intensity in the bottom panel. If and only if there are coding differences between  $A$  and  $B$ , then the slope of the market-level risk curve with respect to marketshare ( $\frac{\partial \bar{r}}{\partial \theta^B}$ ) will be different from zero.

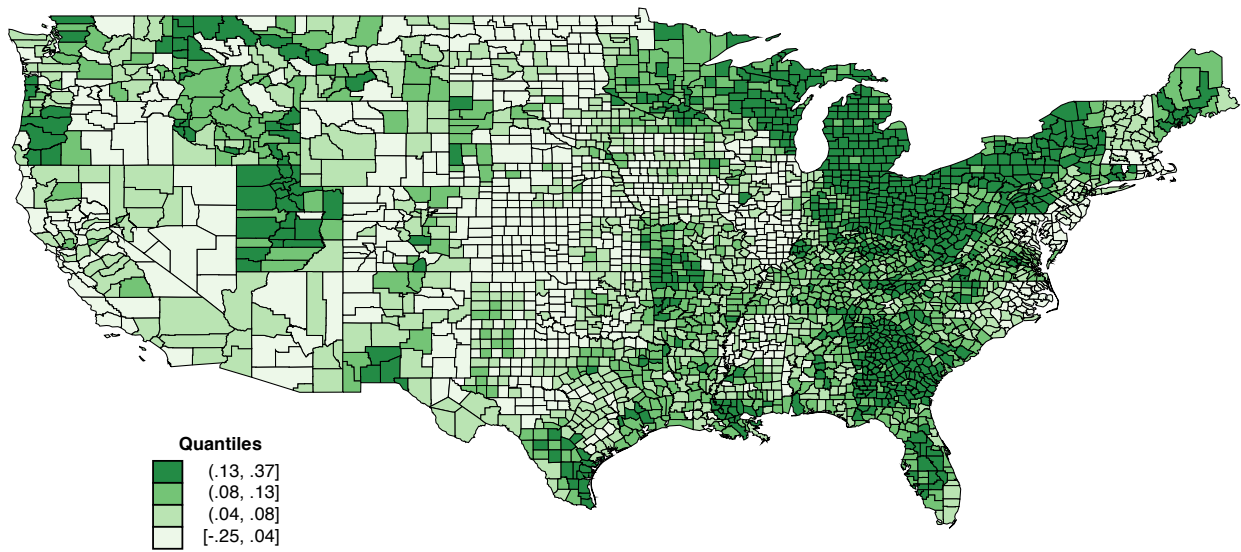


**Figure 3: Growth in Medicare Advantage (MA) Penetration**



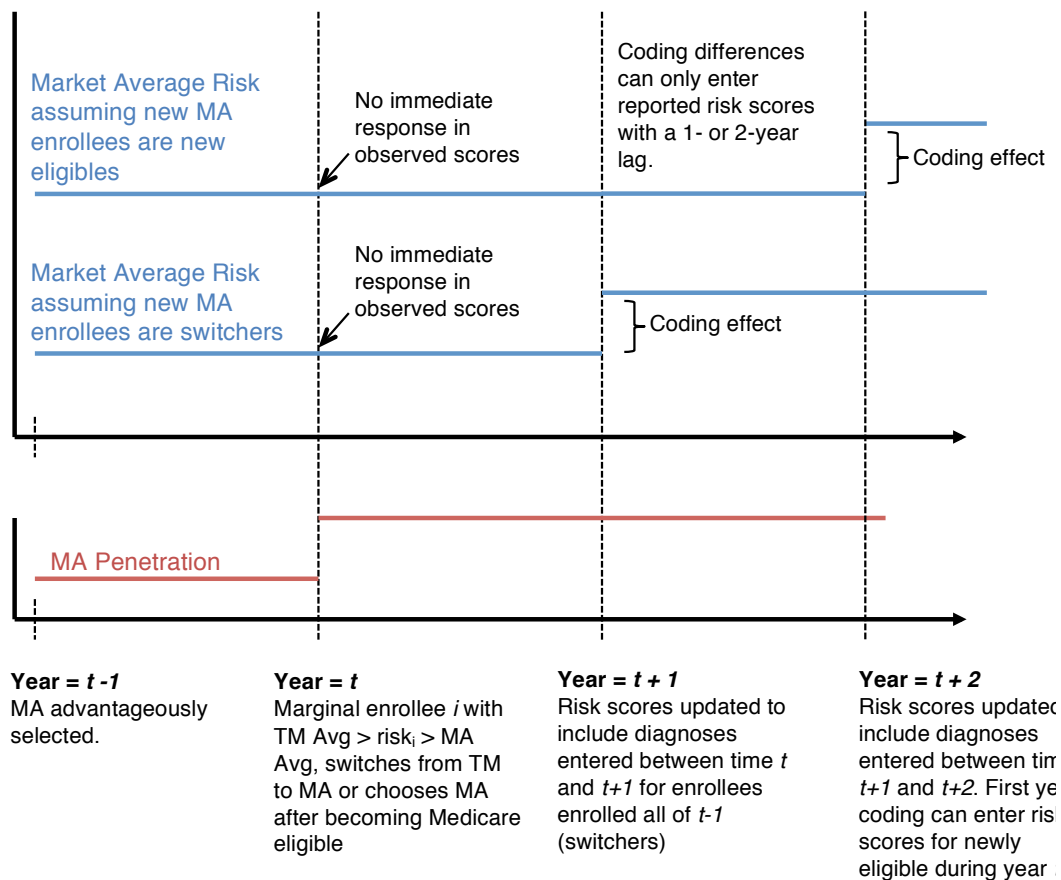
**Note:** The top panel displays national trends in MA penetration, where the unit of observation is the Medicare beneficiary. *Source, pre 2006:* Kaiser Family Foundation, 2013. The bottom panel displays a histogram of within-county changes in penetration from 2006 to 2011, using the main estimation sample. The unit of observation is the county.

**Figure 4:** Geography of Growth in Medicare Advantage (MA), 2006 to 2011



**Note:** Map shows changes by county in MA penetration from the beginning to end of our sample period, 2006 to 2011. Counties are binned and color-coded according to their quantile of changes in penetration. Darker regions indicate larger MA growth.

**Figure 5: Timing Illustration: Coding Effects Occur with a Lag in Medicare**



**Note:** This diagram highlights the timing of changes in market level average risk in response to a change in MA penetration. For the first year in either MA or TM, a switcher carries forward a risk score based on his last year in the other segment. For the newly eligible (those turning 65), demographic risk scores are assigned until there is a full year of enrollment and diagnoses information. Therefore, upcoding effects should not be apparent until a full year following the change in enrollment when the penetration change is due to switchers, and upcoding effects should not be apparent until period  $t+2$  when the penetration change is due new Medicare enrollees.

**Table 1:** Suggestive Evidence: Distributions of Diabetes and Cancer HCCs in TM and MA

Diabetes			
HCC (from least to most severe)	TM	MA	Coefficient
HCC 19 - Diabetes without complication	84.85%	80.40%	0.162
HCC 18 - Diabetes with ophthalmologic	2.82%	2.33%	0.259
HCC 17 - Diabetes with acute complication	0.93%	0.45%	0.339
HCC 16 - Diabetes with neurologic	4.62%	6.14%	0.408
HCC 15 - Diabetes with renal	6.79%	10.67%	0.508
Cancer			
HCC (from least to most severe)	TM	MA	Coefficient
HCC 10 - Breast, Prostate, Colorectal and other cancers and tumors	66.88%	69.06%	0.208
HCC 9 - Lymphatic, head and neck, brain, and other major cancers	11.85%	10.18%	0.794
HCC 8 - Lung, upper digestive tract, and other severe cancers	9.80%	8.85%	1.053
HCC 7 - Metastatic cancer and acute Leukemia	11.47%	11.92%	2.276

**Note:** Table 1A (1B) presents the distributions of diabetes (cancer) HCCs in year 2 for individuals with no diabetes (cancer) diagnosis in year 1 in TM and MA. MA data are derived from the Massachusetts All-Payer Claims Database. Both samples include individuals age 65-88 residing in Massachusetts who were enrolled in the market segment (TM or MA) for all 24 months of the 2011-12 (2010-11 for TM) period. The TM sample does not include dual-eligible beneficiaries. The coefficients are from the 2011 CMS-HCC risk adjustment model. Details on data construction in Appendix [A.2](#).

**Table 2: Summary Statistics**

	Analysis Sample: Balanced Panel of Counties, 2006 to 2011				
	2006		2011		Obs
	Mean	Std. Dev.	Mean	Std. Dev.	
MA penetration (all plan types)	7.1%	9.1%	16.2%	12.0%	3128
Risk (HMO/PPO) plans	3.5%	7.3%	10.5%	10.5%	3128
PFFS plans	2.7%	3.2%	2.7%	3.7%	3128
Employer MA plans	0.7%	2.2%	2.8%	4.3%	3128
Other MA plans	0.2%	1.4%	0.0%	0.2%	3128
MA-Part D Only Penetration	6.5%	9.5%	13.1%	10.8%	3128
MA non-Part D Only Penetration	0.6%	1.7%	3.0%	4.0%	3128
Market Risk Score	1.057	0.084	1.054	0.090	3128
Risk Score in TM	1.064	0.087	1.057	0.089	3128
Risk Score in MA	0.949	0.181	1.032	0.155	3124
Ages within Medicare					
<65	19.8%	6.3%	17.2%	6.2%	3128
65-69	23.5%	3.4%	23.7%	3.1%	3128
70-74	19.2%	1.9%	20.2%	2.5%	3128
75-79	15.9%	2.1%	15.4%	1.8%	3128
≥80	21.6%	4.4%	23.5%	5.0%	3128

**Note:** Table shows county-level summary statistics for the first and last year of the main analysis sample. The sample consists of 3128 counties for which we have a balanced panel of data on penetration and risk scores. MA penetration in the first row is equal to the beneficiary-months spent in Medicare Advantage divided by the total number of total Medicare months spent in the county  $\times$  year. The market risk score is averaged over all Medicare beneficiaries in the county.

**Table 3: Main Results**

	Dependent Variable: County-Level Average Risk Score								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MA penetration $t$ (placebo)	0.016 (0.016)	0.004 (0.020)	0.004 (0.020)	0.017 (0.014)	0.006 (0.018)	0.006 (0.018)	0.006 (0.027)	-0.019 (0.047)	-0.017 (0.047)
MA penetration $t-1$	0.070** (0.011)	0.068** (0.012)	0.066** (0.012)	0.068** (0.011)	0.069** (0.011)	0.067** (0.011)	0.090** (0.021)	0.070** (0.025)	0.069** (0.024)
Main Effects									
County FE	X	X	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X	X	X
Additional Controls									
State X Year Trend		X	X		X	X		X	X
County-Year Demographics			X			X			X
MA non-Part D Penetration				X	X	X			
MA Part D Penetration							X	X	X
Mean of Dep. Var.	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03
Observations	15,640	15,640	15,640	15,640	15,640	15,640	15,640	15,640	15,640

**Note:** Regressions of average reported risk scores in the market (county) on Medicare Advantage (MA) penetration.  $t-1$  indicates penetration in the county in the prior plan year. Observations are county-years. In columns (1) through (3), all variation in MA penetration is used. In columns (4) through (6) only variation in MA-Part D penetration is used, and in columns (7) through (9) only variation in MA non-Part D penetration is used. Observations are county-years. Standard errors in parentheses are clustered at county level. \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table 4:** Placebo Tests: Effects of Contemporaneous Penetration and Leads

	Dependent Variable: County-Level Average Risk Score			
	2007-2011	2007-2010	2007-2009	2008-2011
	(1)	(2)	(3)	(4)
MA penetration t+2 (placebo)			0.045+ (0.023)	
MA penetration t+1 (placebo)		0.016 (0.025)	0.028 (0.058)	
MA penetration t (placebo)	0.004 (0.020)	-0.019 (0.029)	-0.062 (0.072)	0.010 (0.018)
MA penetration t-1	0.066** (0.012)	0.076** (0.018)	0.083** (0.022)	0.045** (0.016)
MA penetration t-2				0.047* (0.023)
Main Effects				
County FE	X	X	X	X
Year FE	X	X	X	X
Additional Controls				
State X Year Trend	X	X	X	X
County-Year Demographics	X	X	X	X
Observations	15,640	12,512	9,384	12,512

**Note:** Regressions of average reported risk scores in the market (county) on Medicare Advantage (MA) penetration.  $t-1$  indicates penetration in the county in the prior plan year; MA Penetration  $t+1$  indicates penetration in the county in one plan year in the future. Observations are county-years. The data include penetration from 2006 through 2011 and market risk from 2007 through 2011. The length of the panel determines the sample size in each column. Standard errors in parentheses are clustered at county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table 5:** Falsification Test: Effects on the Demographic Portion of the Risk Score

	Dependent Variable: Demographic Portion of County-Level Average Risk Score		
	(1)	(2)	(3)
MA penetration t	0.000 (0.002)	0.001 (0.002)	0.001 (0.002)
MA penetration t-1	0.001 (0.002)	0.000 (0.002)	-0.001 (0.002)
Main Effects			
County FE	X	X	X
Year FE	X	X	X
Additional Controls			
State X Year Trend		X	X
County-Year Demographics			X
Mean of Dep. Var.	0.485	0.485	0.485
Observations	15,640	15,640	15,640

**Note:** Regressions of average demographic risk scores in the market (county) on Medicare Advantage (MA) penetration. Demographic risk scores are calculated by the authors using data from Medicare Beneficiary Summary File. *t*-1 indicates penetration in the county in the prior plan year. Observations are county-years. Standard errors in parentheses are clustered at county level.\*  $p < 0.05$ , \*\*  $p < 0.01$ .



**Table 6:** Falsification Test: Effects on Morbidity and Mortality

	Dependent Variable:					
	Mortality over 65			Cancer Incidence over 65		
	(1)	(2)	(3)	(4)	(5)	(6)
MA penetration t	-0.002 (0.002)	0.002 (0.002)	0.002 (0.003)	-0.005 (0.004)	-0.005 (0.005)	-0.005 (0.005)
MA penetration t-1	0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	0.005 (0.004)	0.001 (0.004)	0.003 (0.005)
Main Effects						
County FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Additional Controls						
State X Year Trend		X	X		X	X
County-Year Demographics			X			X
Mean of Dep. Var.	0.048	0.048	0.048	0.023	0.023	0.023
Observations	15,408	15,408	15,408	3,050	3,050	3,050

**Note:** Regressions of average mortality/cancer incidence in the market (county) on Medicare Advantage (MA) penetration. Mortality and cancer incidence data from SEER dataset. Sample size reflects number of counties where SEER measures mortality and cancer incidence. *t*-1 indicates penetration in the county in the prior plan year. Observations are county-years. Standard errors in parentheses are clustered at county level.\*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table 7:** Falsification Test: Effects on Age Distribution

	Fraction $\geq 65$	Conditional on $\geq 65$				
		65-69	70-74	75-79	80-84	85+
	(1)	(2)	(3)	(4)	(5)	(6)
MA penetration $t$	0.003 -(0.004)	0.002 -(0.007)	0.008 -(0.007)	-0.004 -(0.006)	-0.001 -(0.006)	-0.006+ -(0.003)
MA penetration $t-1$	-0.004 -(0.004)	-0.006 -(0.006)	0.019** -(0.006)	-0.006 -(0.007)	-0.003 -(0.006)	-0.004 -(0.004)
Main Effects						
County FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Additional Controls						
State X Year Trend	X	X	X	X	X	X
County-Year Demographics						
Observations	15,640	15,640	15,640	15,640	15,640	15,640

**Note:** Regressions of the population age distribution on Medicare Advantage (MA) penetration. columns 2 through 6 are conditional on  $\geq 65$ . Time  $t-1$  indicates penetration in the county in the prior plan year. Observations are county-years. Standard errors in parentheses are clustered at county level.\*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table 8:** The Role of Insurer-Provider Integration: Effects by Contract Type

		Dependent Variable: County-Level Average Risk Score		
		(1)	(2)	(3)
HMO/PPO/POS penetration	t (placebo)	0.004 -(0.033)	0.019 -(0.037)	0.017 -(0.037)
PFFS penetration	t (placebo)	-0.003 -(0.033)	0.008 -(0.038)	0.008 -(0.037)
Employer MA penetration	t (placebo)	0.004 -(0.012)	-0.026+ -(0.014)	-0.024+ -(0.013)
HMO/PPO/POS penetration	t-1	0.138** -(0.026)	0.096** -(0.027)	0.095** -(0.027)
PFFS penetration	t-1	0.041+ -(0.023)	0.061* -(0.025)	0.057* -(0.025)
Employer MA penetration	t-1	0.041** -(0.011)	0.043** -(0.012)	0.043** -(0.012)
Main Effects				
County FE		X	X	X
Year FE		X	X	X
Additional Controls				
State X Year Trend			X	X
County-Year Demographics				X
Observations		15,640	15,640	15,640

**Note:** Regressions of average reported risk scores in the market (county) on Medicare Advantage (MA) penetration by MA plan type. *t*-1 indicates penetration of indicated plan type in the county in the prior plan year. Observations are county-years. Observations are county-years. Standard errors in parentheses are clustered at county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

## APPENDIX

### A.1 Background on MA Risk-Adjusted Payments

Medicare Advantage (MA) insurance plans are given monthly capitated payments for each enrolled Medicare beneficiary. The bases of these county-level payments are tied to historical Traditional Medicare (TM) costs in the county. County-based payments were originally intended to capture the cost of enrolling the “national average beneficiary” in the Traditional Medicare program in the county, though Congress has made many *ad-hoc* adjustments over time.

Before 2004, there was relatively minimal risk adjustment of capitation payments, which relied primarily on demographics. In 2004, CMS began transitioning to risk adjustment based on diagnoses obtained during inpatient hospital stays and outpatient encounters. By 2007, diagnosis-based risk adjustment was fully phased-in. During our study period (2006-2011), risk adjusted capitation payments were equal to  $R_{ic}^j = \phi_c^j \cdot x_{ic} \Lambda$ , where  $i$  indexes beneficiaries,  $j$  indexes plans, and  $c$  indexes counties (markets). The basis  $\phi^j$  was approximately equal to the county “benchmark”  $\phi_c$ , though  $\phi_c^j$  could vary across plans within the same county.

$\phi$  could vary within counties because since 2006, MA plans have been required to submit bids to CMS, which are compared to the uniform county benchmark  $\phi_c$ . If the bid is below the county benchmark set by the regulator, the plan receives 75% of the difference between the bid and benchmark, which the plan folds back into its premium and benefits as a “rebate” to beneficiaries. Importantly for our purposes, this 75% is still paid out by CMS into the MA program, so that what matters for the implicit subsidies to MA that we calculate here is the capitation payment to plans inclusive of any “rebate.” Therefore, we abstract from discussion of the details of how the rebate is allocated.

### A.2 Data Used in Section 2

For the analysis in Section 2, we obtained data on Medicare Advantage claims from the new Massachusetts All-Payer Dataset (Mass APCD). The Mass APCD includes the universe of health insurance claims for individuals receiving medical services in the state of Massachusetts. Payers, along with third-party claims administrators and pharmacy benefit managers, report all claims to the state of Massachusetts. These claims are then aggregated into a large, comprehensive dataset. To identify individuals, we use an individual ID created by the state generated using Social Security numbers.

We also obtained data on Traditional Medicare Fee-For-Service claims from CMS. We use the Master Beneficiary Summary File along with the Research Identifiable Inpatient, Outpatient, and Carrier claims files for 2010 and 2011. While for the MA sample we use data from 2011-12, for the TM sample we use data from 2010-11 due to the 2012 data being unavailable at the time we performed the analysis. We eliminate all dual eligibles from the TM data to be consistent with the Mass APCD data which does not include Medicaid claims.

For the MA sample, we isolated all individuals continuously enrolled in one of the four largest Medicare Advantage plans for all 24 months of the 2011-12 time period.<sup>38</sup> Analogously, in the TM sample we isolated all individuals continuously enrolled in TM for all 24 months of the 2010-11 time period. We restricted both samples to individuals with ages between 65 and 88, inclusive.<sup>39</sup>

<sup>38</sup>We restrict the sample to these four large Medicare Advantage plans because the variables used in the Mass APCD to indicate that an individual is enrolled in Medicare Advantage exhibit some inconsistencies when compared to publicly available Medicare Advantage enrollment data from CMS. To ensure that we are observing Medicare Advantage claims from Medicare Advantage insurers, we isolate Medicare Advantage plans using payer names, product IDs, the variables provided in the dataset indicating whether the product is a Medicare Advantage product, and the age distribution of enrollees for each product.

<sup>39</sup>We don’t include individuals with ages greater than 88 because age is top-coded at 88 in the Mass APCD data.

For the analysis, we isolate all individuals in each sample with no diabetes diagnosis in year 1 but at least one diabetes diagnosis in year 2.<sup>40</sup> Isolating individuals with no diabetes diagnosis in year 1 allows us to be more confident that the individuals in the MA and TM samples are of similar health status.

---

<sup>40</sup>We define diabetes diagnoses as the set of diagnoses that map to one of the five diabetes HCCs shown in Table 1

**Table A1:** Extended Placebo Tests: Effects of Contemporaneous Penetration and Leads

	Dependent Variable: County-Level Average Risk Score						
	available panel years: 2007-2011	2007-2010	2007-2009	2008-2011	2008-2010	2008-2009	2009-2011
	(1)	(2)	(3)	(1)	(2)	(3)	(3)
MA penetration t+2 (placebo)			0.045+ (0.023)			0.029 (0.038)	
MA penetration t+1 (placebo)		0.016 (0.025)	0.028 (0.058)		-0.005 (0.016)	-0.023 (0.043)	
MA penetration t (placebo)	0.004 (0.020)	-0.019 (0.029)	-0.062 (0.072)	0.010 (0.018)	0.007 (0.026)	-0.019 (0.091)	0.014 (0.017)
MA penetration t-1	0.066** (0.012)	0.076** (0.018)	0.083** (0.022)	0.045** (0.016)	0.036 (0.023)	0.024 (0.039)	0.040 (0.034)
MA penetration t-2				0.047* (0.023)	0.058* (0.024)	0.051 (0.041)	0.050 (0.032)
MA penetration t-3							(0.025) 0.025
Main Effects							
County FE	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
Additional Controls							
State X Year Trend	X	X	X	X	X	X	X
County-Year Demographics	X	X	X	X	X	X	X
Observations	15,640	12,512	9,384	12,512	9,384	6,256	9,384

**Note:** Regressions of average reported risk scores in the market (county) on Medicare Advantage (MA) penetration. *t*-1 indicates penetration in the county in the prior plan year; MA Penetration *t*+1 indicates penetration in the county in one plan year in the future. Observations are county-years. The data include penetration from 2006 through 2011 and market risk from 2007 through 2011. The length of the panel determines the sample size in each column. Standard errors in parentheses are clustered at county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .