

# Upcoding or Selection?

## Evidence from Medicare on Squishy Risk Adjustment\*

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

October 21, 2015

### Abstract

Upcoding—manipulation of patient diagnoses in order to game payment systems—has gained significant attention following the introduction of risk adjustment into US insurance markets. We provide new evidence that enrollees in private Medicare plans generate 6% to 16% higher diagnosis-based risk scores than they would generate under fee-for-service Medicare, where diagnoses do not affect payments. Our estimates imply upcoding generates billions of dollars in excess public spending annually and significant consumer choice distortions. We show that coding intensity increases with vertical integration, reflecting a principal-agent problem faced by insurers, who desire more intense coding from the physicians with whom they contract.

---

\*We thank Colleen Carey, Joshua Gottlieb, and Amanda Kowalski for serving as discussants, as well as seminar participants at the 2014 Annual Health Economics Conference, the 2014 American Society of Health Economists Meeting, the BU/Harvard/MIT Health Economics Seminar, Boston University, Emory University, Harvard Medical School, the NBER Public Economics Meeting 2015, the University of Illinois at Chicago, RTI, the Southeastern Health Economics Study Group, and the University of Texas at Austin for useful comments. We also thank Chris Afendulis, Marika Cabral, Vilsa Curto, David Cutler, Francesco Decarolis, Liran Einav, Randy Ellis, Keith Ericson, Amy Finkelstein, Austin Frakt, Craig Garthwaite, Jonathan Gruber, Jonathan Kolstad, Tom McGuire, Hannah Neprash, Joe Newhouse, and Daria Pelech for assistance obtaining data and useful conversations. Layton gratefully acknowledges financial support from the National Institute of Mental Health (T32-019733). Geruso gratefully acknowledges financial support from the Robert Wood Johnson Foundation and from grants 5 R24 HD042849 and 5 T32 HD007081 awarded to the Population Research Center at the University of Texas at Austin by the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

<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: [layton@hcp.med.harvard.edu](mailto:layton@hcp.med.harvard.edu)

# 1 Introduction

Diagnosis-based subsidies have become an increasingly important regulatory tool in US health insurance markets. Between 2003 and 2014, the number of consumers enrolled in a market where an insurer's payment is based on the consumer's diagnosed health conditions increased from almost zero to about 50 million, including in Medicare, Medicaid, and state and federal Health Insurance Exchanges. These diagnosis-based payments to insurers are typically known as risk adjustment. By compensating insurers for enrolling high expected-cost consumers, risk adjustment weakens insurer incentives to engage in cream-skimming—that is, inefficiently distorting insurance product characteristics to attract lower-cost enrollees as in [Rothschild and Stiglitz \(1976\)](#).<sup>1</sup> It also works to solve the [Akerlof \(1971\)](#) problem of high-cost enrollees driving up contract prices, leading to the kind of market unravelling that has attracted significant attention in the empirical public finance literature (e.g. [Einav, Finkelstein and Cullen, 2010](#) and [Hackmann, Kolstad and Kowalski, 2015](#)).

The intuition underlying risk adjustment is straightforward, but the mechanism assumes that a regulator can objectively measure each consumer's health state. In practice in health insurance markets, regulators infer an enrollee's health state from the diagnoses reported by physicians during their encounters with the enrollee. This diagnosis information is aggregated into a risk score on which insurer payments are based. The mechanism generates a connection between physician coding behavior and insurer payments, providing a strong incentive for insurers to influence physicians to “upcode” diagnosis information in the patient's record. For instance, insurers can pay physicians on the basis of the codes they assign, rather than on the basis of visits and procedures performed. The extent to which insurers are successful in acting on this incentive in practice is of considerable policy, industry, and popular interest, in particular because in many settings upcoding implies a clear cost to taxpayers.<sup>2</sup>

Nonetheless, relatively little is known about the extent of upcoding or its implications: The few recent studies examining the distortionary impacts of risk adjustment ([Brown et al., 2014](#), [Carey, 2014](#), [Geruso and McGuire, 2015](#), and [Einav et al., 2015](#)) have all taken diagnosis coding as fixed, rather than as an endogenous outcome determined by physician and insurer strategic behavior. In

---

<sup>1</sup>For instance, in Medicare Advantage a diagnosis of the condition Diabetes with Acute Complications generates a payment for the private insurer that is incrementally larger by about \$3,390 per year. This amount is set by the regulator to be equal to the average incremental cost incurred by individuals diagnosed with Diabetes with Acute Complications in the traditional fee-for-service Medicare program.

<sup>2</sup>See, for example, [CMS, 2010](#); [Government Accountability Office, 2013](#); [Kronick and Welch, 2014](#); [Schulte, 2014](#).

contrast, in this paper we show that endogenous diagnosis coding is an empirically important phenomenon leading to billions in annual overpayments by the federal government, as well as significant consumer choice distortions.

We begin by constructing a simple framework that assumes a risk score is generated by an enrollee  $\times$  insurer match, and therefore would vary for the same individual depending on her choice of insurance plan. We show that if risk scores vary across insurers for the same consumer, then even when risk adjustment is perfectly successful in counteracting both the [Rothschild and Stiglitz \(1976\)](#) and [Akerlof \(1971\)](#) selection inefficiencies, it introduces a new distortion by differentially subsidizing plans with higher coding intensity. This differential payment for intensive coding has two effects: First, in a partially privatized public insurance program like Medicare, upcoding increases the program's cost to taxpayers as more intensely coded patients generate higher government subsidies (and there is no mechanism to enforce budget neutrality). Second, because the differential payment implicitly creates a voucher that is larger when consumers choose a plan with higher coding intensity, consumer choices can be inefficiently tilted toward the plans that code most intensely.<sup>3</sup> This choice distortion has not been previously noted, though we show that it operates and is likely to be economically important in any risk-adjusted health insurance market, including Medicare, many state Medicaid programs, and the Health Insurance Exchanges created by the Affordable Care Act (ACA).

We investigate the empirical importance of upcoding in the context of Medicare. For hospital and physician coverage, Medicare beneficiaries can choose between a traditional fee-for-service (FFS) option and enrolling with a private insurer through Medicare Advantage (MA). In the FFS system, most reimbursement is independent of recorded diagnoses. Payments to private MA plans, however, are capitated with diagnosis-based risk adjustment. Although the incentive for MA plans to code intensely is strong, a plan's response to this incentive depends on its ability to influence the providers that assign the codes.<sup>4</sup> Thus, whether and to what extent coding differs between the MA and FFS segments of the market is an empirical question.

The key challenge in identifying coding intensity differences between FFS and MA, or within the MA market segment among competing insurers, is that upcoding estimates are potentially con-

---

<sup>3</sup>We show below that choices are distorted irrespective of why coding intensity differs across plans or market segments. Even if the higher coding intensity is valued by consumers—for example, because higher intensity is associated with better continuity of care—the differential reimbursement is nonetheless distortionary.

<sup>4</sup>In addition, if generating higher risk scores requires more (costly) contact with patients, then MA plans face a countervailing incentive because the plan is the residual claimant on any dollars not spent on patient care.

founded by adverse selection. An insurer might report an enrollee population with higher-than-average risk scores either because the consumers who choose the insurer's plan are in worse health (selection) or because for the same individuals, the insurer's coding practices result in higher risk scores (upcoding). We develop an approach to separately identify selection and coding differences in equilibrium. The core insight of our research design is that if the same individual would generate a different risk score under two insurers and if we observe an exogenous shift in the market shares of the two insurers, then we should also observe changes in the *market-level* average of reported risk scores. Such a pattern could not be generated or rationalized by selection, because selection can affect only the sorting of risk types across insurers within the market, not the overall market-level distribution of reported risk scores.<sup>5</sup> A key advantage of our strategy is that the data requirements are minimal, and it could be easily implemented in future assessments of coding in Health Insurance Exchanges or state Medicaid programs.<sup>6</sup> Our focus on empirically disentangling upcoding from selection distinguishes our study from prior, policy-oriented work investigating upcoding in the context of Medicare (e.g. [Kronick and Welch, 2014](#)).<sup>7</sup>

We exploit large and geographically diverse increases in 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.<sup>8</sup> We simultaneously exploit an institutional feature of the MA program that causes risk scores to be based on prior year diagnoses. This yields sharp predictions about the timing of effects relative to changing market penetration in a difference-in-differences framework. 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 0.64 percentage point increase in the reported average risk score in a county.

---

<sup>5</sup>Our strategy is analogous to that of [Chetty, Friedman and Rockoff \(2014\)](#), who identify teacher value-added (purged of student selection) via changes in the composition of teaching staff within a grade over time. [Chetty, Friedman and Rockoff \(2014\)](#) analogously avoid biases introduced by endogenous student sorting across teachers by examining changes in outcomes aggregated up to the grade level, rather than at the teacher level.

<sup>6</sup>The key advantages here are that market-level average risk scores and plan enrollment data are sufficient for analysis and that the strategy does not rely on a natural experiment that changes how particular codes are reimbursed. Only exogenous shifts in plan market shares are needed for identification.

<sup>7</sup>[Kronick and Welch \(2014\)](#) provide evidence that risk scores have grown more rapidly over time in MA relative to FFS. An important difference from our analysis is that that study's strategy, which compares the growth rate of risk scores in the FFS population to the growth rate of risk scores in the MA population, would not be robust to selection on health. Further, by focusing on differences in growth rates, the [Kronick and Welch \(2014\)](#) strategy cannot estimate the parameter of interest here—the difference in risk scores and implied payments for a consumer choosing MA versus FFS. Nonetheless, it is the strongest prior evidence that MA codes intensively relative to FFS.

<sup>8</sup>The current risk adjustment system was implemented in MA in 2004 and fully phased in by 2007. Data on risk scores in MA are available from the regulator beginning in 2006.

This implies that MA plans generate risk scores for their enrollees that are on average 6.4% larger than what those same enrollees would have generated under FFS. Further, we show that the size of this coding difference is related to the vertical relationship between insurers and providers. For example, health maintenance organizations (HMOs) have the highest coding intensity of all plan types operating in the Medicare Advantage market. Ownership is also important: Fully vertically integrated (i.e., physician owned) plans generate 16% higher risk scores for the same patients compared to FFS, nearly triple the effect of non-integrated plans.

To complement our main identification strategy at the market level, we also provide individual-level evidence for a small sample of Massachusetts residents. We track risk scores within consumers as they transition from an employer or individual-market commercial plan to Medicare at the age 65 eligibility threshold. In the individual data, consumers choosing FFS Medicare have similar risk scores and diagnoses in their employer plans at age 64 and in FFS Medicare at age 65, but consumers choosing MA show a spike in reported risk scores and diagnoses in the year of transition from their employer plans to an MA plan. This entirely separate identification strategy based on the Medicare eligibility threshold confirms the size of our estimates from the main analysis and allows us to examine the specific conditions that account for the coding differences.

These empirical findings have specific implications for Medicare as well as broader implications for the regulation of private insurance markets. Medicare is the costliest public health insurance program in the world and makes up a significant fraction of US government spending. Absent a coding correction, our estimates imply excess payments of around \$10.5 billion to Medicare Advantage plans annually, or about \$640 per MA enrollee per year.<sup>9</sup> These public finance impacts are large because the coding difference of 6.4% that we estimate is large: a 6.4% increase in the average risk score is equivalent to 6% of all consumers in a market becoming paraplegic, 11% of all consumers developing Parkinson's disease, or 39% becoming diabetics. While these effects would be implausibly large if they reflected rapid changes to actual population health, they are plausible when viewed as reflecting only endogenous coding behavior. In fact, our average estimates align closely with regulators' stated beliefs and policy actions regarding upcoding (CMS, 2010; Government Accountability Office, 2013).<sup>10</sup>

---

<sup>9</sup>In 2010, toward the end of our study period, CMS began deflating MA risk scores by 3.41% due to concerns about upcoding. Our results indicate that the 2010 deflation adjustment is both too small and fails to account for large coding differences across insurance contract types.

<sup>10</sup>The Government Accountability Office has consistently argued that CMS should deflate risk scores by 5-7%. CMS

In the context of the broader literature concerned with the public finance of Medicare, the magnitude of the overpayments due to upcoding is striking. For example, [Decarolis \(2015\)](#) investigates how Medicare Part D insurers manipulate bids to game payment formulas and drive up payments; [Ho, Hogan and Scott Morton \(2014\)](#) estimate excess public spending arising from consumers' inattention to health plan choice and insurers' endogenous responses to that inattention; and [Brown et al. \(2014\)](#) estimate the increase in excess payments to MA plans due to uncompensated favorable selection following the implementation of risk adjustment. While each of these phenomena are important—each with billions in public spending at stake—the potential excess spending due to unchecked upcoding is even larger.<sup>11,12</sup> To provide further context for the size of the effects that we estimate, we draw on estimates of demand response from the prior literature on MA. These estimates imply that completely removing the hidden subsidy due to upcoding would reduce the size of the MA market by 17% to 33%, relative to a counterfactual in which CMS made no adjustment.<sup>13</sup>

We view our results as addressing an important gap in the literature on adverse selection and the public finance of healthcare. Risk adjustment is the most widely implemented regulatory response to adverse selection. A few recent studies, including [Curto et al. \(2014\)](#) and [Einav and Levin \(2014\)](#), have begun to recognize the potential importance of upcoding, but the empirical evidence is underdeveloped. The most closely related prior work on coding has shown that patients' reported diagnoses vary with the local practice style of physicians ([Song et al., 2010](#)) and that coding responds to

---

began deflating MA risk scores in 2010, toward the end of our study period. The deflator was set at 3.41% in 2010; was increased to 4.91% in 2014; and is set to increase again to 5.91% in 2015, consistent with our findings.

<sup>11</sup>The excess public spending implied by unaddressed upcoding (\$640 per MA enrollee per year) is twice as large as the [Brown et al. \(2014\)](#) estimate of the increase in excess payments to MA plans due to uncompensated favorable selection. It is also more than three times the size of the excess government payments in Medicare Part D that [Ho, Hogan and Scott Morton \(2014\)](#) estimate arise from consumers' inattention to health plan choice. Similar to [Decarolis \(2015\)](#), [Ho, Hogan and Scott Morton \(2014\)](#), and [Brown et al. \(2014\)](#), our identifying variation is not suited to performing a full welfare analysis, and our focus is on the public spending implications. Nonetheless, we discuss the channels by which welfare is affected in Section 7.

<sup>12</sup>Our findings also contribute to the growing policy literature on the broader welfare impacts of the MA program. In addition to the benefits of expanding choice, one popular argument in favor of MA is that it might create important spillover effects for FFS 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, Chernen and Robbins \(2013\)](#) find that the expansion of MA resulted in lower hospital costs in FFS Medicare. Our findings indicate that these benefits of privatized Medicare do not come without costs. Any positive spillovers should be balanced alongside the additional costs (the deadweight loss of taxation plus welfare losses due to choice distortions) of upcoding in MA.

<sup>13</sup>These estimates compare two counterfactuals: one in which CMS does not adjust risk scores to account for upcoding, and another in which CMS fully adjusts for the mean coding difference of 6.4%. In 2010, toward the end of our study period, CMS began deflating risk scores by 3.41%, removing about half of the overpayment. The same estimates of demand response from the literature imply that increasing the deflation factor from 3.41% to the full 6.4% we estimate would have reduced the size of the MA market by 8% to 15% relative to the actual size in 2010.

changes in how particular codes are reimbursed by FFS Medicare for inpatient hospital stays ([Dafny, 2005](#); [Sacarny, 2014](#)). Ours is the first study to model the implications of differential coding patterns across insurers and to provide empirical evidence of the determinants such differences. As we discuss below, these findings are directly relevant for understanding consumer choice distortions due to coding in the new state and federal Health Insurance Exchanges created by the ACA.

Finally, our results provide a rare insight into the insurer-provider relationship. Because diagnosis codes ultimately originate from provider visits, insurers face a principal-agent problem in contracting with physicians. Our finding that coding intensity varies significantly according to the contractual relationship between the physician and the insurer suggests that the cost of aligning physician incentives with insurer objectives may be significantly lower in vertically integrated firms. These results connect to a long literature concerned with the internal organization of firms and the application of these ideas to the healthcare industry (e.g., [Gaynor, Rebitzer and Taylor, 2004](#) and [Frakt, Pizer and Feldman, 2013](#)), as well as to the study of the intrinsic ([Kolstad, 2013](#)) and extrinsic ([Clemens and Gottlieb, 2014](#)) motivations of physicians. Our results also represent the first direct evidence of which we are aware that vertical integration between insurers and providers facilitates the “gaming” of health insurance payment systems. However, these results likewise suggest that strong insurer-provider contracts may also facilitate other, more socially beneficial, objectives, including quality improvements through pay-for-performance incentives targeted at the level of the insurer. This is an issue of significant policy and research interest (e.g., [Fisher et al., 2012](#); [Frakt and Mayes, 2012](#); [Frandsen and Rebitzer, 2014](#)), but as [Gaynor, Ho and Town \(2015\)](#) describe in their recent review, it is an area in which there is relatively little empirical evidence.

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, and the implications for public spending. In Section 3, we explain our strategy for estimating upcoding in the presence of selection. In Section 4, we discuss our data and empirical setting. In sections 5 and 6, we present results, and in Section 7 we discuss implications for government overpayments and consumer choice distortions. Section 8 concludes.



## 2 Model of Risk Adjustment with Endogenous Coding

### 2.1 Risk Adjustment Background

We begin by briefly describing the functioning of a risk-adjusted payment system in a regulated private insurance market. Plans receive payment from a regulator for each individual they enroll, which supplements or replaces premiums paid by the enrollee. The net payment  $R$  after risk adjustment for enrolling individual  $i$  is equal to the individual's risk score,  $r_i$ , multiplied by some benchmark amount,  $\phi$ , set by the regulator:  $R_i = \phi \cdot r_i$ .<sup>14</sup> The regulator distributes risk adjustment payments from a fund, or enforces transfers across plans.<sup>15</sup> The risk score itself is calculated by multiplying a vector of risk adjusters,  $x_i$ , by a vector of risk adjustment coefficients,  $\Lambda$ . Net payments are therefore  $R_i = \phi \cdot x_i \Lambda$ . By compensating the insurer for an enrollee's expected cost, risk adjustment makes all potential enrollees appear equally profitable to the insurer, in principle, since all enrollees have the same *net* expected cost. This removes incentives to distort contracts to attract lower-cost enrollees, as in [Rothschild and Stiglitz \(1976\)](#). Because risk adjustment does not compensate for realized costs, insurers still bear all risk and remain the residual claimants on lower patient healthcare spending.

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 based on diagnosis codes contained in health insurance claims. In Medicare, as well as the federal Health Insurance Exchanges, these indicators 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 each condition on the insurer's expected costs, as estimated by the regulator in a regression of total spending on the vector  $x_i$  in some reference population. Coefficients  $\Lambda$  are normalized by the regulator so that the average risk score is equal to one in the relevant population. One 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, diagnosed medical conditions are properties of individuals, not individual  $\times$  plan matches.

---

<sup>14</sup>The benchmark payment can be equal to the average premium paid in the full population of enrollees, as in the ACA Exchanges, or some statutory amount, as in Medicare Advantage.

<sup>15</sup>The fund can be financed via tax revenues or via fees assessed to health plans by the regulator. In Medicare Advantage, the fund is financed by tax revenues, while in the Exchanges the fund is financed by health plan fees.



## 2.2 Upcoding Defined

To explore the implications of upcoding for public spending, we relax the assumption that risk scores are invariant to an enrollee’s plan choice by allowing the reported conditions for individual  $i$  to vary by plan  $j$ . More specifically, we model risk scores as generated in the following way:<sup>16</sup>

$$r_{ij} = \hat{r}_i + \alpha(\gamma_j, \psi_j), \quad (1)$$

where  $\alpha(\gamma, \psi)$  describes a plan-specific coding factor that is a function of plan characteristics.<sup>17</sup> We define characteristics  $\gamma$  as not valued by consumers, despite their impact on risk scores. For example, insurers may implement automated retrospective chart reviews to identify missing codes within the insurer’s systems, with no effect on the consumer’s experience. Characteristics  $\psi$  are valued by consumers and could include features like home health visits, which impact recorded diagnoses and patient utility via utilization.

From here forward, we define differential coding intensity—or relative “upcoding”—between plans as *any* (valued or non-valued) differences across plans that would result in the same individual being assigned different risk scores in different plans, or  $\alpha(\gamma_j, \psi_j) \neq \alpha(\gamma_k, \psi_k)$ . This formulation does not require that coding differences arise only as an insurer response to the financial incentive to code intensely; nor does it assume that coding activities are independent of enrollee utility.

This plan-dependent definition of the risk score generates an analogous plan-dependent definition of the risk-adjusted payment:

$$R_{ij} = \phi \cdot r_{ij} = \phi \cdot (\hat{r}_i + \alpha(\gamma_j, \psi_j)). \quad (2)$$

It directly follows that if plan  $j$  codes more intensively than  $k$  in the sense of generating a higher  $\alpha$ , then  $j$  would receive a larger payment than  $k$  for enrolling the same individual. We refer to the difference  $\phi r_{ij} - \phi r_{ik} = \phi(\alpha(\gamma_j, \psi_j) - \alpha(\gamma_k, \psi_k))$  as the differential voucher. It is measured in dollars.

To understand the potential impact of this differential voucher on public spending, it is useful to apply the idea to our empirical setting. For hospital and physician coverage, Medicare benefi-

<sup>16</sup>Section 2.3 provides institutional details about how risk scores are generated in practice.

<sup>17</sup>This assumption requires that coding differences across plans be the same for all enrollees. Here, we maintain this assumption for tractability. In the next section, we relax this assumption and show that it is not necessary for many empirical applications of the model.

ciaries can choose between using the public fee-for-service option, where individual providers are reimbursed based on procedures performed, and enrolling with a private insurance plan under the Medicare Advantage option. Under Medicare Advantage, insurers are paid by Medicare on a risk-adjusted, capitated basis as in Eq. 2. Private insurers then make payments to their providers under various arrangements.

For the MA market segment only, the regulator attempts to set benchmarks ( $\phi$ ) and risk adjustment coefficients ( $\Lambda$ ) so that the payment to the Medicare Advantage insurer for consumer  $i$  would just equal the total cost of reimbursing providers for procedures if  $i$  were enrolled under fee-for-service Medicare.<sup>18</sup> In order to do this, the risk adjustment coefficients  $\Lambda$  are generated using fee-for-service data on total costs and conditions ( $x_i$ ) and thus reflect the relationship between costs and conditions under fee-for-service Medicare. Therefore, if an individual is assigned more condition codes under a Medicare Advantage plan, she will generate a larger insurer payment relative to her counterfactual FFS cost. Specifically, government spending on Medicare Advantage is higher by the amount  $\phi(r_{i,MA} - r_{i,FFS}) = \phi(\alpha(\gamma_{MA}, \psi_{MA}) - \alpha(\gamma_{FFS}, \psi_{FFS}))$ . Summing over the entire population, the extra government spending due to coding differences between FFS and MA is equal to:

$$\sum_{i=1}^N (\phi(\alpha(\gamma_{MA}, \psi_{MA}) - \alpha(\gamma_{FFS}, \psi_{FFS})) \cdot 1[MA_i]), \quad (3)$$

where  $1[MA_i]$  is an indicator for choosing a Medicare Advantage plan. If coding is identical in MA and FFS, the differential voucher is zero, and government spending on individual  $i$  is unaffected by plan choice.

The source of the coding differences (valued versus non-valued plan characteristics) is irrelevant in this calculation once the difference in the two  $\alpha$  terms is pinned down. Further, we have not defined an objectively “correct” level of coding; only relative differences matter. In our empirical setting, this means that it does not matter whether physicians billing under FFS Medicare pay too little attention to coding or whether MA insurers pay too much attention to coding. As long as  $\alpha(\gamma_{MA}, \psi_{MA}) > \alpha(\gamma_{FFS}, \psi_{FFS})$ , government spending will be higher in MA than FFS.

Beyond the budgetary effect of the overpayment—which, given the \$600 billion/year size of Medicare, is important in itself—the differential voucher also has a distortionary impact on the con-

<sup>18</sup>In practice, during this period the process by which  $\phi$  is set is more complicated and involves urban and rural “floors” that do not allow  $\phi$  to go below statutory minimum values.

sumer’s choices. By awarding a voucher with a higher dollar value to consumers who choose an MA plan, the subsidy to coding intensity has the effect of tilting consumer choices toward the more intensely coded Medicare Advantage market segment (and within MA, toward the most intensely coded MA plans). The intuition here is straightforward: As would be true for taxes or subsidies applied in any market, a subsidy that varies in size according to the product selected will distort prices away from marginal costs and be welfare decreasing, relative to a first best. In our setting, this is true even if higher coding intensity is associated with plans delivering higher quality.<sup>19</sup> We formalize the intuition in Section 7, where we discuss the relevance of our findings for other risk-adjusted markets including the Exchanges. In the Exchanges, the institutions are such that the subsidy to upcoding is budget neutral from the government’s perspective. We show that even under budget-neutrality, upcoding has this distortionary effect on consumer choices.

## 2.3 Coding in Practice

In most markets with risk adjustment, regulators recognize the potential for influencing diagnosis codes and attempt to respond by placing restrictions on which diagnosis codes can be used to determine an individual’s risk score. Typically, the basis for all valid diagnosis codes is documentation from a face-to-face encounter between the provider and the patient. During an encounter 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, including diagnosis codes, that is sent to the insurer for payment.<sup>20</sup> The insurer pays the claims and over time aggregates all of the diagnoses associated with an enrollee to generate a risk score on which the payment from the regulator is based.

In Figure A1, we outline the various mechanisms insurers employ to affect diagnosis coding, and in turn risk scoring.<sup>21</sup> We exclude any mechanisms that involve illegal action on the part of insurers.<sup>22</sup> First, and before any patient-provider interaction, insurers can structure contracts with physician groups such that the payment to the group is a function of the risk-adjusted payment that the insurer itself receives from the regulator. This directly passes through coding incentives to the

---

<sup>19</sup>Intuitively, if consumers value quality, they don’t need a differential subsidy to induce them to make the appropriate selection. We formalize this logic in Section 7.

<sup>20</sup>Traditionally, the diagnoses were included on the claim to provide justification for the service for which the provider was billing the insurer.

<sup>21</sup>Insights in the figure come from investigative reporting by the Center for Public Integrity, statements by CMS, and our own discussions with MA insurers and physician groups.

<sup>22</sup>While fraud is a known problem in health insurance markets, coding differences can easily arise without any explicitly illegal actions on the part of the insurer.

physician groups. Insurers may also choose to selectively contract with providers who code more aggressively. Additionally, the insurer can influence coding during the medical exam by providing tools to the physician that pre-populate his notes with information on prior-year diagnoses for the patient. Since risk adjustment in many settings, including MA, is based solely on the diagnoses from a single plan year, this increases the probability that diagnoses, once added, are retained indefinitely. Insurers also routinely provide training to the *physician's* billing staff on how to assign codes to ensure the coding is consistent with the insurer's financial incentives. Finally, even after claims and codes are submitted to the insurer for an encounter, the insurer may automatically or manually review claims, notes, and charts and either add new codes based on physician notes that were not originally translated to the claims or request a change to the coding in claims by the physician's billing staff.<sup>23</sup>

In addition to 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 enrollees to 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 undiscovered. Further, if an insurer observes that an enrollee whose expected risk score is high based on medical history has not visited a physician in the current plan year, the insurer can directly intervene by proactively contacting the enrollee or sending 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 and relatively low cost. There is substantial anecdotal evidence and lawsuits related to such behavior in Medicare Advantage,<sup>24</sup> and regulators have expressed serious concern that such visits primarily serve to inflate risk scores.<sup>25</sup>

None of these insurer activities take place in FFS because providers under the traditional system

---

<sup>23</sup>Insurers use various software tools to scan patient medical records for diagnoses that were not recorded in originally submitted claims, but nonetheless meet the statutory eligibility requirements and can therefore be added as valid codes.

<sup>24</sup>See, for example, [Schulte \(2015\)](#).

<sup>25</sup>In a 2014 statement, CMS 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 that 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."

are paid directly by the government, the basis of these payments is procedures, not diagnoses.<sup>26,27</sup> This difference in incentive structure between FFS and MA makes Medicare a natural setting for studying the empirical importance of upcoding, especially given that the risk adjustment models used in Medicare were designed with the objective of minimizing the potential for gaming (Pope et al., 2004).

### 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 coding differs 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 a non-chosen plan during the same plan year is unobservable.

Our solution to the identification problem is to focus on market-level risk. Consider a large geographic market in which the population distribution of *actual* health conditions is stationary. In such a setting, market-level *reported* risk scores could nonetheless change if market shares shift between plans with higher and lower coding intensity. Our strategy is related to that of Chetty, Friedman and Rockoff (2014), which identifies teacher value-added (purged of student selection) via changes in the composition of teaching staff within a grade over time, using outcomes aggregated up to the grade level, rather than at the teacher level.<sup>28</sup>

#### 3.1 Graphical Intuition

Figure 1 provides the graphical intuition for this idea. We depict two plans, or market segments, labeled  $j$  and  $k$ . They are intended to align with TM and MA, respectively. All consumers choose

---

<sup>26</sup>Under FFS, hospitals are compensated for inpatient visits via the diagnosis-related groups (DRG) payment system, in which inpatient stays are reimbursed partially based on inpatient diagnoses and partially based on procedures. It is nonetheless plausible that overall coding intensity in FFS and MA differs significantly. For one, the set of diagnoses compensated under the inpatient DRG payment system differs from that of the MA HCC payment system. In addition, the majority of FFS claims are established in the outpatient setting, in which physician reimbursement depends on procedures, not diagnoses.

<sup>27</sup>In FFS, diagnoses are instead used for the purpose of providing justification for the services for which the providers are requesting reimbursement.

<sup>28</sup>There the selection problem avoided is endogenous student sorting across teachers. An even closer analog of our exact method to the educational context would be for use in 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.

either  $j$  or  $k$ . Plan  $k$  is assumed to be advantageously selected on risk scores, so that the risk score of the marginal enrollee is higher than that of the average enrollee.<sup>29</sup> The top panel shows three curves: the average risk in  $j$  ( $\bar{r}_j$ ), the average risk in  $k$  ( $\bar{r}_k$ ), and the average risk of all enrollees in the market ( $\bar{r}$ ).

In the top panel of Figure 1, we plot the baseline case of no coding differences across plans. The market share of  $k$ , denoted by  $\theta^k$ , increases along the horizontal axis. Average risk in  $k$  is low at low levels of  $\theta^k$  because the few beneficiaries selecting into  $k$  are the lowest risk. As long as there is no coding difference between  $j$  and  $k$ , the market-level risk ( $\bar{r}$ ), which is averaged over enrollees in both plans, is constant in  $\theta^k$ . This is because reshuffling enrollees across plans within a market does not affect the market-level distribution of underlying health conditions.

The bottom panel of Figure 1 incorporates differential coding: For any individual, plan  $k$  is assumed to assign a risk score higher than that assigned under  $j$  by some constant factor as in Eq. 9. For reference, the dashed line in the figure represents the counterfactual average risk that plan  $k$  enrollees would have been assigned under  $j$ 's coding intensity,  $\bar{r}_k^j$ . The key insight is that in the bottom panel where coding intensity differs, the slope of market-level risk  $\bar{r}$  with respect to  $k$ 's market share ( $\theta^k$ ) is non-zero. Intuitively,  $\frac{\partial \bar{r}}{\partial \theta^k}$  reveals upcoding because the marginal consumer switching from  $j$  to  $k$  increases  $\theta^k$  and simultaneously increases the average reported risk ( $\bar{r}$ ) in the market by moving to a plan that assigns her a higher score. Although the bottom panel of Figure 1 depicts the empirically relevant case in which the advantageously selected plan or market segment is more intensely coded, we show next that the same intuition applies regardless of the presence or direction of selection.<sup>30</sup>

### 3.2 Model

We now generalize this graphical analysis to allow for consumer preferences, consumer risk scores, and plan characteristics that generate arbitrary patterns of selection. We also allow for a more general representation of coding differences across plans.

Continue to assume that all consumers choose between two plans or market segments, labeled  $j$

<sup>29</sup>Note that this figure does not describe selection on costs net of risk adjustment, but rather selection on risk scores. This is because our goal here is to distinguish between *risk score* differences due to coding and *risk score* differences due to selection. If selection existed only along net costs (and not risk scores), then estimating coding intensity differences would be trivial. One could directly compare the means of risk scores across plans.

<sup>30</sup>For illustration, in Appendix Figure A2, we depict a case in which the advantageously selected plan codes less intensely, a case where coding differences exist absent any selection on the risk score, and a case in which selection is both nonlinear and non-monotonic.

and  $k$ . As in Section 2,  $r_{ik} - r_{ij}$  represents  $i$ 's differential risk score.  $\theta^k$  is defined as above, and  $\mathbb{1}[k_i(\theta^k)]$  is an indicator function for choosing  $k$ , which expresses, for any level of  $k$ 's market share, the plan choice of consumer  $i$ . Then, the average risk score in the market can be expressed as

$$\bar{r} = \frac{1}{N} \sum \left( r_{ij} + \mathbb{1}[k_i(\theta^k)](r_{ik} - r_{ij}) \right). \quad (4)$$

The top panel of Figure 1 illustrates that when coding is homogenous across insurers, the market-level average risk does not vary with market shares. To see that this holds generally under any pattern of selection between plans, note that if coding is identical in plans  $j$  and  $k$ , then  $r_{ik} = r_{ij}$  for every enrollee, implying  $\bar{r} = \frac{1}{N} \sum (r_{ij})$  and  $\frac{\partial \bar{r}}{\partial \theta^k} = 0$ .

The bottom panel of Figure 1 suggests that if coding differs between plans ( $r_{ij} \neq r_{ik}$ ), then  $\frac{\partial \bar{r}}{\partial \theta^k} \neq 0$ . Under the assumption from Section 2 that an individual's risk score  $r_{ik}$  is composed of a plan-independent individual risk component  $\hat{r}_i$  plus a plan-dependent component  $\alpha(\gamma_j, \psi_j)$ , the slope of the market average risk curve  $\bar{r}$  exactly pins down the coding difference between  $j$  and  $k$ :

$$\frac{\partial \bar{r}}{\partial \theta^k} = \alpha(\gamma_k, \psi_k) - \alpha(\gamma_j, \psi_j). \quad (5)$$

With additive separability of the individual and plan-specific components of the risk score, Equation 5 holds for any distribution of risks and for any form of selection across plans.<sup>31</sup> In Appendix A.2, we allow for heterogeneity in coding at the individual  $\times$  plan level ( $r_{ij} = \hat{r}_i + \alpha(\gamma_j, \psi_j) + \epsilon_{ij}$ ) so that individuals are heterogeneous in the extent to which their risk scores differ across plans. We prove that in this case, Eq. 5 still identifies the mean difference in risk scores across plans ( $\alpha(\gamma_k, \psi_k) - \alpha(\gamma_j, \psi_j)$ ). The proof requires only that any heterogeneity in coding at the individual  $\times$  plan level is orthogonal to  $\theta^k$ .

Eq. 5 implies that in the typical risk adjustment scheme in which risk scores are normed to one, if plan  $k$  generates risk scores 10% of the mean higher than what  $j$  would generate for the same consumers, then the slope of the market-level risk curve with respect to  $k$ 's market share would be 0.10. If these assumptions fail to hold (i.e., if coding heterogeneity at the individual  $\times$  plan level is systematically related to market penetration), then the market-level average risk curve will be nonlinear in

---

<sup>31</sup>Proof: Recall that  $r_{ij} = \hat{r}_i + \alpha(\gamma_j, \psi_j)$ . Then  $\frac{\partial \bar{r}}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{1}{N} \sum \left( \hat{r}_i + \alpha_j + \mathbb{1}[k_i(\theta)](\alpha_k - \alpha_j) \right) = (\alpha_k - \alpha_j) \cdot \frac{\partial}{\partial \theta} \frac{1}{N} \sum \mathbb{1}[k_i(\theta)] = \alpha_k - \alpha_j$ . This makes no assumption on the distribution of  $\hat{r}_i$  or on joint distribution of risks and preferences that generate the selection curves  $\bar{r}_j(\theta)$  and  $\bar{r}_k(\theta)$ .



$\theta^k$  because the marginal and average coding differences will not be equal. In that case,  $\frac{\partial \bar{r}}{\partial \theta^k}$  is a local approximation that identifies the coding difference among the marginal enrollees. Our aggregate, market-level data (which do not include individual risk scores) cannot be used to identify heterogeneity across individual enrollees, though we discuss in Section 7 how this kind of heterogeneity would affect the interpretation of our estimates of the public spending implications.

## 4 Setting and Empirical Framework

We apply the identification insights from Section 3 to examine coding differences between Medicare Advantage (MA) plans and the traditional fee-for-service (FFS) program. We begin with a brief 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 FFS program administered by the federal government or coverage through a private MA plan. MA plans are attractive to Medicare beneficiaries because, compared to the traditional system, these plans 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 utilize 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. In 2004, CMS began transitioning from risk adjustment that was based primarily on demographics to risk adjustment based on diagnoses obtained during inpatient hospital stays and outpatient encounters. By 2007, diagnosis-based risk adjustment was fully phased-in.

As described in Section 2, the capitation payment is the product of the benchmark rate  $\phi_c$ , which varies across counties  $c$ , and a person-specific risk score  $r_{ij}$ , which may be endogenous to the choice of plans ( $j$ ). From the consumer's perspective, all Medicare-eligible consumers in a county face the same menu of MA plan options at the same prices. Historically, county benchmarks have been set

to capture the cost of covering the “national average beneficiary” in the FFS program in that county, though Congress has made many *ad-hoc* adjustments over time.<sup>32</sup> CMS sets risk adjustment coefficients nationally using claims data from FFS.

## 4.2 Data

Estimating the slope  $\frac{\partial \bar{r}}{\partial \theta^{MA}}$  from Figure 1 requires observing market-level risk scores at varying levels of MA penetration. We obtained yearly county-level averages of risk scores and MA enrollment by plan type from CMS for 2006 through 2011.<sup>33</sup> MA enrollment is defined as enrollment in any MA plan type, including managed care plans like Health Maintenance Organizations (HMOs) and Preferred Provider Organizations (PPOs), private fee-for-service (PFFS) plans, and employer MA plans.<sup>34</sup> Average risk scores within the MA and FFS market segments are weighted by the fraction of the year each beneficiary was enrolled in the segment. We define MA penetration ( $\theta^{MA}$ ) as the fraction of all beneficiary-months of a county-year spent in an MA plan. For most of our analysis, we collapse all MA plans together and consider the markets as divided between the MA and FFS segments, though we also analyze the partial effects of increased penetration among various subsets of MA plan types.

All analysis of risk scores is conducted at the level of market averages, as the regulator does not generally release individual-specific risk adjustment data for MA plans.<sup>35</sup> 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 to construct county-level averages of the demographic (age and gender) component of risk scores, which we use in a falsification test.<sup>36</sup>

Table 1 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

<sup>32</sup>In practice, benchmarks can vary somewhat by plan. See Appendix A.1 for full details.

<sup>33</sup>Similar data are unavailable before 2006, since diagnosis-based risk scores were not previously generated by the regulator.

<sup>34</sup>We exclude only enrollees in the Program of All-inclusive Care for the Elderly (PACE) plans.

<sup>35</sup>CMS has not traditionally provided researchers with individual-level risk scores for MA enrollees (two exceptions are Brown et al. (2014) and Curto et al. (2014)). A strength of our identification strategy, which could easily be applied in other settings like Medicaid Managed Care and Health Insurance Exchanges, is that it does not require individual-level data.

<sup>36</sup>The regulator’s algorithm specifies that the demographic components ( $r_i^A$ ) and diagnostic components ( $r_{ij}^{Dx}$ ) of individual risk scores are additively separable, which implies that the county averages are also additively separable:  $\bar{r} = \frac{1}{N_c} \sum_{i \in I_c} (r_i^A + r_{ij}^{Dx}) = \bar{r}^A + \bar{r}^{Dx}$ .

through the last year for which data are available, 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 an overall market mean of approximately 1.0, are lower within MA than within FFS, implying that MA selects healthier enrollees.<sup>37</sup> Table 1 also shows the dramatic increase in MA penetration over our sample period, which comprises one part of our identifying variation.

## 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 MA. Additionally, MA 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 2, we put this timing in historical context, charting the doubling of MA penetration nationally between 2005 and 2011. The bottom panel of the figure shows that within-county penetration changes were positive in almost all cases, though the size of these changes varied widely. Figure A3 shows that this MA penetration growth was not limited to certain regions or to urban areas. Each county is shaded according to its quartile of penetration changes.

Our main identification strategy relies on year-to-year variation in penetration within geographic markets to trace the slope of the market average risk curve,  $\frac{\partial \bar{r}}{\partial \theta^{MA}}$ . The identifying assumption in our difference-in-differences framework is that year-to-year growth in MA enrollment within counties did not track year-to-year variation in the county’s actual population-level health. The assumption is plausible, given that county population health, reflected in the incidence of chronic conditions used in risk scoring, such as diabetes and cancer, is unlikely to change sharply year-to-year. In contrast, *reported* risk can change instantaneously due to coding differences when a large fraction of the Medicare population moves to MA. Additionally, the within-county changes in MA penetration

---

<sup>37</sup>For estimation, we normalize the national average to be exactly 1.0 in each year.

appear as large shocks rather than slow shifts in enrollment trends, suggesting supply-side rather than demand-side factors are responsible for the variation in MA penetration we exploit. Further, we can test the assumption of no correlated underlying health trends for a variety of independently observable demographic, morbidity, and mortality outcomes at the county level.

#### 4.3.2 Timing

We also exploit an institutional feature of how risk scores are calculated in MA 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, upcoding should be apparent only with a lag relative to penetration changes.

We illustrate the timing in Figure 3. The individual's risk score that is reported and 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 FFS during open enrollment in January of a given year, the risk score for her entire first year in MA will be based on diagnoses she received while in FFS during the prior calendar 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 newly-eligible Medicare beneficiaries choosing to enroll in MA. In order for an MA enrollee to be assigned a diagnosis-based risk score, CMS requires the enrollee to have accumulated a full calendar year of diagnoses. This restriction causes all new MA enrollees to be assigned demographic risk scores during their first calendar year of MA enrollment. Additionally, many individuals first enroll in MA when they become eligible for Medicare on their 65th birthday. This results in most new MA enrollees joining MA partway through a calendar year, causing them to also have an incomplete set of diagnoses from their first calendar year of enrollment. These enrollees receive a demographic risk score during their first *and second* years of MA enrollment. This is illustrated in Figure 3, and it implies that if coding intensity is higher in MA, changes in MA penetration due to newly-eligible 65-year-olds should affect reported coding with up to a two-year lag. We exploit these timing features below.

#### 4.4 Econometric Framework

The slope of the market-level average risk score with respect to MA penetration identifies coding intensity in MA relative to FFS. 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 difference-in-differences models of the form:

$$\bar{r}_{sct} = \gamma_c + \gamma_t + \sum_{\tau \in T} \beta_{\tau} \cdot \theta_{sct}^{MA} + f(X_{sct}) + \epsilon_{sct}, \quad (6)$$

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  $\gamma_t$ , so that coefficient estimates for  $\beta$  are identified within counties across time and scaled to the level of the continuous treatment variable ( $\theta^{MA}$ ).  $X_{sct}$  is a vector of time-varying county characteristics described in more detail below. The subscript  $\tau$  in the summation indicates the timing of the penetration variable relative to the timing of the reported risk score. This specification allows flexibility in identifying post-penetration change effects as well as pre-trends (placebos). 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 FFS enrollees and after the second year of MA enrollment for newly-eligible beneficiaries. Because of the short panel nature of the data—the regulator's data start in 2006 and our extract ends in 2011—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 FFS. Under our additive coding intensity assumption,  $\beta_{t-1} + \beta_{t-2}$  is exactly equal to  $\alpha(\gamma_{MA}, \psi_{MA}) - \alpha(\gamma_{FFS}, \psi_{FFS})$ , which is the difference between the risk scores that the same individual would generate under the two systems.

We include the placebo regressor that captures contemporaneous effects of MA enrollment changes

$(\theta_{sc, \tau=t}^{MA})$  in all specifications. The coefficient on the placebo reveals whether counties were differentially trending in the dependent variable prior to when risk scores could have plausibly been affected by upcoding. The contemporaneous effect of penetration changes on market-level risk reflected in  $\beta_t$  should be zero.

## 5 Results

We begin by presenting the results on coding that include all Medicare Advantage (MA) plan types.<sup>38</sup> After reporting on a series of falsification and placebo tests in support of our identifying assumption, we examine how coding differences vary according to the level of integration between the insurer and providers.

### 5.1 Main Results

Table 2 reports our main results. 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 difference-in-differences coefficient indicates that the market-level average risk score in a county increases by about 0.07—approximately one standard deviation—as lagged MA penetration increases from 0% to 100%. Because risk scores are scaled to have a mean of one, this implies that an individual’s risk score in MA is about 7% higher than it would have been under fee-for-service (FFS) Medicare. In column 2, we add linear state time trends, and in column 3, we add time-varying controls for county demographics.<sup>39</sup> Across specifications, the coefficient on lagged MA penetration is stable.

To put the size of these coding effects in context, a 0.07 increase in market-level risk is equivalent to 7% of all consumers in the market becoming paraplegic, 12% of all consumers developing Parkinson’s disease, or 43% becoming diabetics. If, contrary to our identifying assumption, these estimates were capturing a spurious correlation between actual changes in underlying health in the local market and changes to MA penetration, it would require that there were large negative shocks

<sup>38</sup>In Appendix A.3, we perform an analogous exercise examining how the within-FFS and within-MA average risk scores vary with MA penetration to provide evidence on selection. We find weak evidence consistent with the common finding of (compensated) advantageous selection into MA on the risk score (e.g., Newhouse et al., 2012), though estimates are imprecise.

<sup>39</sup>These controls consist of 18 variables that capture the fraction of Medicare beneficiaries in the county-year in five-year age bins from 0 to 85 and 85+.

to population health that closely tracked enrollment changes. This would imply that (with high frequency) insurers' contract design and pricing was changed to become more attractive at the times and in the places where population health was rapidly deteriorating.<sup>40</sup> Further, we provide some direct evidence below that population health was not trending differentially with MA penetration, using alternative health and mortality data that is external to the reports from MA plans.

The effects we estimate would be implausibly large if they were due to true changes in underlying population health. However, if the estimates reflect only differential coding, as we claim, then these parameter estimates are closely consistent with widely held beliefs about coding in MA. The Government Accountability Office has expressed concerns that coding differences between MA and FFS are in the range of 5% to 7% ([Government Accountability Office, 2013](#)), consistent with our estimates. In 2010 toward the end of our data period, CMS began applying a deflation factor to MA risk scores when determining payments to private plans in the MA program, to correct for the assumed upcoding by private MA plans. Originally set at 3.41% in 2010, that deflation is now slated to increase to 5.91% for 2015. Our econometric estimates (6.4%) confirm the beliefs implied by these policy actions, though, as we show below in Section 5.4, the mean MA-FFS coding difference masks important heterogeneity not previously considered by regulators. We show that the degree of coding intensity is closely tied to plan-provider integration.

As a robustness check, we also estimate versions of the main regressions that isolate different sources of year-to-year variation in MA enrollment. MA-Part D plans combined MA with the new Medicare prescription drug benefit introduced in 2006. In Appendix Table A2, we re-estimate our regressions controlling for changes in MA penetration in any plan type other than MA-Part D. This identifies effects using only the penetration growth directly attributable to growth in the MA-Part D market. We alternatively control for the complement of this variation: changes in penetration by MA-Part D plans. This identifies estimates using only changes in MA penetration arising from enrollment in plans that did not offer a Part D drug benefit. All results are closely consistent with Table 2, which uses all within-county across-time variation.

---

<sup>40</sup>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 also seems plausible, as the literature suggests that if there is a relationship between demand for MA and health status, it is that as health deteriorates demand for MA decreases rather than increases ([Brown et al., 2014](#); [Newhouse et al., 2012](#)).



## 5.2 Placebo/Parallel Trends Tests

The coefficient estimates for contemporaneous MA penetration in Table 2, which are close to zero and insignificant across all specifications, constitute a placebo test. Because we observe no contemporaneous correlation between MA penetration changes and changes in risk scores, these coefficients imply that the health of the population was not drifting in a way that was spuriously correlated with changes in penetration. Instead, effects appear only with a lag, consistent with the institutions described in Figure 3. 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. Due to the length of time diagnosis-based risk adjustment has existed in Medicare, the 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 the three years from 2007 to 2009. Nonetheless, in columns 1 through 3 of Table 3, we repeat the main analysis with additional leads, under the intuition that significant coefficients on contemporaneous effects or leads would provide evidence against the parallel trends assumption.

The column headers in Table 3 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  $\theta_{t+2}^{MA}$ ) show no consistent evidence that contemporaneous or future values of MA penetration are correlated with market-level changes in time  $t$  risk scores. Because true population characteristics, especially the prevalence of the chronic conditions that determine risk scores, tend to change gradually rather than discretely, the large and precisely timed response with a lag of at least one year is more consistent with a mechanical coding effect than a change in true population health.

As discussed in the context of Figure 3, switchers from FFS to MA carry forward their old risk scores for one plan-year, but newly-eligible consumers aging into Medicare and choosing MA will not have risk scores based on diagnoses that were assigned by their MA plan until after two plan years.<sup>41</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 in column 4 indicate a cumulative

---

<sup>41</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 may ratchet up the longer an individual is enrolled in MA. Even for switchers from FFS, this could result in positive coefficients for more than a single lag of MA penetration.

upcoding effect of 8.7% ( $=4.1+4.6$ ) after two years. In the short panel, we are limited in estimating the effects of longer lags or leads with any precision. Nonetheless, we report on an extended set of leads and lags in Appendix Table A3, which supports the robustness of our findings.

### 5.3 Falsification Tests

In Tables 4, 5, and A4, 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 not plausibly affected by upcoding. In particular, we focus on county demographics, mortality, and morbidity.

Table 4 replicates the specifications in columns 1 through 3 of Table 2, but with the demographic portion of the risk score 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 because CMS retrieves this information from Social Security data. The coefficients, which are near zero and insignificant in all specifications, show no impact of MA penetration, consistent with the mechanism we describe in which enrollees are assigned more, or more severe, medical conditions to drive up risk scores.<sup>42</sup>

In Table 5, we test whether changes in MA penetration are correlated with mortality or morbidity. Columns 1 through 3 show the relationship between changes in a county's mortality rate among county residents age 65 and older and changes in MA penetration. For morbidity, finding illness data that is not potentially contaminated by the insurers' coding behavior is challenging. We use cancer incidence data from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute, which operates an independent data collection enterprise and does not rely on insurer claims to determine diagnoses. Columns 4 through 6 show the relationship between cancer incidence among county residents age 65 and older 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. Across both outcomes in Table 5 and for all specifications, coefficients on contemporaneous and lagged MA penetration are consistently close to zero and statistically insignificant.

---

<sup>42</sup>An additional implication of the results in Table 4 (also consistent with our identifying assumption) 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.

Finally, we test whether changes in MA penetration are correlated with changes in the county's Medicare enrollee age distribution in Table A4. The estimates show no consistent evidence of a systematic relationship between MA penetration and the Medicare enrollee age distribution. In sum, each falsification test supports our identifying assumption of no correlation between MA penetration and actual underlying population health or demographics.

## 5.4 Insurer-Provider Integration

Because diagnoses originate with providers rather than insurers, insurers face an agency problem regarding coding. Plans that are physician owned, selectively contract via physician networks, or follow managed care models (i.e., HMOs and PPOs) may have more tools available for influencing provider coding patterns. For example, our conversations with MA insurers and physician groups indicated that vertically integrated plans often pay their physicians (or physician groups) partly or wholly as a function of the risk score that physicians' diagnoses generate. In such cases, within large physician groups, leadership may use monthly "conferences" to display the risk scores generated by individual physicians and place workplace pressure on low-scoring physicians to bring scores into line with the group. Additionally, insurers may be able to use the threat of removal from their networks to generate adherence to their coding objectives. Integration, broadly defined by the closeness of insurers and providers, could therefore influence a plan's capacity to affect coding.

To investigate this possibility, in Table 6 we separate the effects of market share increases among HMO, PPO, and private fee-for-service (PFFS) plans. HMOs may be the most likely to exhibit integration, followed by PPOs. PFFS plans are fundamentally different. During most of our sample period, PFFS plans did not have provider networks. Instead, PFFS plans reimbursed Medicare providers based on 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 to influence diagnoses recorded within the physician's practice. In particular, PFFS insurers could not arrange a contract with providers that directly rewarded intensive coding, and PFFS insurers would not be likely to train a physician's billing staff on coding. PFFS plans could, nonetheless, affect the probability of diagnoses via consumer incentives for contact with physicians: PFFS plans routinely set lower copays for routine and specialist visits than beneficiaries faced under FFS. PFFS plans could also utilize home visits from providers employed by the insurer to capture codes that otherwise may have been missed.

As in the main analysis, the coefficients of interest in Table 6 are on lagged penetration, while contemporaneous penetration coefficients comprise our standard placebo (parallel trends) test.<sup>43</sup> Point estimates in the table show that the strongest coding intensity is associated with managed care plans generally, and HMOs in particular. Risk scores in HMO plans are around 10% higher than they would have been for the same Medicare beneficiaries enrolled in FFS. PPO coding intensity is around 7% higher than FFS. Risk scores in PFFS plans, while intensely coded relative to FFS, exhibit relatively smaller effects.

In Table 7, we report on a complementary analysis that classifies MA plans according to whether the plan was physician-owned, using data collected by [Frakt, Pizer and Feldman \(2013\)](#). We describe these data in Appendix A.4. The analysis uses physician ownership as an alternative definition of insurer-provider integration. Table 7 yields a similar insight as Table 6: Physician-owned MA plans display coding intensity that is larger than the overall mean among MA plans. Physician ownership is associated with risk scores that are about 16% higher than in FFS Medicare, while the average among all other MA plans is a 6% coding difference.

The results in Tables 6 and 7 show that the overall average upcoding factor masks significant heterogeneity across plan types. This evidence suggests that the costs of aligning physician and insurer incentives may decline significantly with vertical integration. The issue of vertical relationships in healthcare markets is a topic of considerable research interest (e.g., [Gaynor, Rebitzer and Taylor, 2004](#)) and practical importance, but as [Gaynor, Ho and Town \(2015\)](#) describe in their recent review, examination of vertical integration has generally suffered from a “a paucity of empirical evidence.” Here we provide the first evidence that vertical integration facilitates gaming of the regulatory system. However, these results likewise indicate that strong insurer-provider contracts may also facilitate other, more socially beneficial, objectives, including quality improvements through “pay-for-performance” initiatives aimed at the level of the insurer. Such mechanisms, including those underlying Accountable Care Organizations, have attracted significant attention (e.g., [Fisher et al., 2012](#); [Frakt and Mayes, 2012](#); [Frandsen and Rebitzer, 2014](#)), but here as well, evidence has been sparse.

Although the internal organization of the MA plan strongly predicts coding intensity, local adoption of electronic health records (EHRs) appears not to play a significant role. We investigated this

---

<sup>43</sup>These regressions also separately control for penetration by employer-sponsored MA plans and for penetration by the remaining plan types. The MA program includes various specialized alternative plan types, serving a small minority of the Medicare market. These include Cost Plans, Special Needs Plans, and other temporary CMS demonstration plans.

possibility using data on adoption of EHR by office-based physicians available from the Office of the National Coordinator for Health Information Technology in the Department of Health and Human Services. The exercise is described in detail in Appendix Section A.5, and the results are displayed in Table A5. Interactions between lagged penetration and an indicator for high EHR adoption by physician offices in the local market yield coefficients very close to zero, though the standard errors do not rule out small effects.

## 6 Individual-Level Evidence

We next turn to a smaller, individual-level dataset that includes diagnoses. We track how reported chronic conditions and risk scores change within the same consumers when 64-year-olds age into Medicare and choose either FFS or MA. This allows us to (i) demonstrate the robustness of our key empirical results to an entirely different identification strategy, (ii) generate estimates that fully capture any practice-style spillovers in the longer-run equilibrium, and (iii) investigate the diagnosis margins along which upcoding occurs.

### 6.1 Data: Massachusetts All-Payer Claims

In the prior literature, it had not been possible to track the diagnoses of individuals aging into Medicare. Insurer claims data are proprietary, and even conditional on obtaining insurer data, there has generally been no way to link the same individual in different insurers' databases in any market. However, as part of its regulatory activity, the state of Massachusetts has recently begun collecting claims files from private insurers in a way that makes possible this kind of linking.

We use the Massachusetts All-Payer Claims Dataset (APCD) for 2011 and 2012. For 2011 and 2012 only, the APCD includes an individual identifier that allows us to follow consumers across years and health plans as they change insurance status. These data cover all private insurer claims data, including Medicare Advantage, Medigap, employer, and individual-market commercial insurers. Therefore, we can observe a consumer in her employer plan at age 64 and then again in her private MA plan at age 65. FFS claims to Medicare are not directly captured as they are exempted from the regulator's reporting requirements.

To indirectly identify claims belonging to FFS enrollees, we follow an approach developed in

Song and Wallace (2015), and use data from private Medigap plans. For FFS enrollees with supplemental Medigap coverage, Medigap pays some fraction of almost every FFS claim, creating a duplicate record of the information in the FFS claim sent to Medicare.<sup>44</sup> Nationally, about 25% of 65-year old FFS enrollees have a supplemental Medigap policy. Conditional on the FFS enrollee having a Medigap plan, we observe a complete record of all the diagnosis information needed to construct a risk score. To the extent that our sample of observable FFS enrollees is a good proxy for the full FFS population in Massachusetts with respect to changes in coding at age 65, we can estimate the differential change in diagnoses at 65 among MA enrollees relative to FFS enrollees.<sup>45</sup>

We focus on two groups of consumers in the data: all individuals who join an MA plan within one month of their 65th birthday and all individuals who join a Medigap plan within one month of their 65th birthday.<sup>46</sup> We limit the sample to individuals with at least six months of data before and after joining MA or Medigap. Our final sample includes 4,724 Medigap enrollees and 1,347 MA enrollees. We use diagnoses from the claims data to generate two risk scores for each individual based on the diagnosed conditions pre- and post-Medicare enrollment. The levels of our calculated risk scores are lower than in the main analysis above because these data capture less than a full year of claims and diagnosis experience.<sup>47</sup> Risk scores are calculated according to the same Medicare Advantage HCC model regardless of the plan type in which the consumer is enrolled (i.e., employer, individual market, FFS, or MA).<sup>48</sup> These risk scores do not share the lagged property (see Figure 3) of

---

<sup>44</sup>The only claims that Medigap does not pay any part of are hospital readmissions and lab claims (paid in full by FFS). Our analysis assumes these types of claims contain no relevant diagnoses that are not also recorded on another claim for the beneficiary. For hospital readmissions, it is unlikely that the new admission will include relevant diagnoses of a chronic condition that did not appear in a prior admission. Lab claims do not typically include diagnoses. When they do, they are likely to be duplicated in the corresponding physician claims, making them unnecessary for calculating risk scores.

<sup>45</sup>Note that this requirement is weaker than requiring that levels of risk scores are similar between FFS enrollees with and without Medigap. FFS enrollees with Medigap and FFS enrollees without Medigap are unlikely to experience different *changes* in coding upon enrollment in Medicare. This is because Medigap plans, unlike MA plans, solely reimburse cost-sharing for services and do not engage in care management or steer patients to a particular set of providers. Nonetheless, a mechanism by which Medigap could influence coding in principle is by increasing an enrollee's utilization of health care (via a demand response to net price). This could lead to more provider-patient encounters during which codes could be obtained. Though we expect such effects are ignorably small, such a phenomenon would imply our estimates of FFS-MA coding differences below are biased toward zero: This type of demand response to Medigap would imply that the increase in FFS risk scores at age 65 in our sample overstates the average change in risk scores in FFS overall. This implies that our estimates of the differential change in coding in MA relative to FFS would be a lower bound of the true differential change.

<sup>46</sup>In each group, we remove everyone who is not continuously enrolled in MA/Medigap after her 65th birthday and everyone who is not continuously enrolled in a (non-Medicare) employer or commercial plan prior to her 65th birthday.

<sup>47</sup>Because the data are limited to the 2011 and 2012 plan years, the only enrollees for whom we have close to a full year of data in the pre- and post-65 plan year are those with a January birthday.

<sup>48</sup>For each individual, we construct the longest possible pre- and post- periods given the individual's enrollment date and a restriction that the pre- and post- periods include the same months from 2011 and 2012. For example, if an individual enrolled in MA/Medigap in July 2011, her pre-period would consist of January-June 2011, and her post period would consist of January-June 2012. However, if an individual enrolled in MA/Medigap in February 2012, her pre-period would consist of February-December 2011 and her post period would consist of February-December 2012.

the scores from the administrative data used in Sections 4 and 5 as we calculate the scores ourselves based on current-year diagnoses.

## 6.2 Risk Scores across the Age 65 Threshold

In Figure 4, we show how risk scores change differentially for consumers entering MA versus consumers entering FFS. The figure plots means of reported chronic conditions and risk scores at age 64 and 65, stratified by whether the beneficiary enters MA or FFS at age 65.<sup>49</sup> The intuition for this strategy is straightforward: If underlying health in the two groups of Medicare beneficiaries follows a similar age trend within a very short window around age 65, then any differential change in risk scores at age 65 between the two groups will be evidence of differential coding in MA vs. FFS.

Evidence that coding intensity is higher in MA than in FFS is apparent by comparing the changes across the age 65 threshold. Figure 4 shows that the *reported* health status of MA enrollees appears to immediately and dramatically worsen at the time of enrollment, relative to the reported trajectory of FFS enrollees. This holds for the probability of being coded with at least one chronic condition that forms the basis for MA risk scores (panel A), the count of these chronic conditions (panel B), and the risk score (panel C). We also note that, consistent with many studies of advantageous selection into MA (Newhouse et al., 2012; Brown et al., 2014), all three panels of Figure 4 suggest that individuals who choose to enroll in MA are unconditionally healthier at age 64 than individuals who choose to enroll in FFS, as reflected in their pre-Medicare diagnoses.

The corresponding difference-in-differences regression results are displayed in Table 8 in specifications that allow for person fixed effects. Column headers indicate the dependent variables. The results in columns 1, 3, and 5 control for individual covariates such as gender, the region of Massachusetts in which the individual resides, and the month the individual enrolled in Medicare. The results in columns 2, 4, and 6 include individual fixed effects to control for any unobserved individual characteristics, including the pre-65 commercial plan choice.

Consistent with the figure, all regressions show a differentially large increase in risk scores and diagnoses at age 65 for new MA enrollees. Columns 1 and 2 indicate that at age 65 the risk scores of new MA enrollees increase by 0.044 more than the increase among new FFS enrollees. These point estimates align very closely with the results presented above in our main dataset and empirical

---

<sup>49</sup>All statistics are based on risk scores that are adjusted for an individual's month of enrollment (and thus the length of the pre- and post- periods), gender, and the region of Massachusetts in which she resides.



strategy: The coefficient on  $\text{Post-65} \times \text{Selected MA}$  is equal to 8% of the mean ( $= \frac{.044}{.550}$ ). Columns 3 through 6 show that this differential increase in risk scores is due to a differential increase in the number of chronic conditions of 0.123 (22% of the mean) and in the probability of being coded for at least one chronic condition of 0.06 (18% of the mean). These results are precisely estimated and robust to the inclusion of individual fixed effects.

These results support and complement the findings of our main analysis, though it is important to understand the limitation of the individual-level analysis. The constrained time dimension of the available data (2011 and 2012 only) combined with a research design that exploits a difference around consumers' 65th birthdays means that we observe pre- and post-65 diagnoses for less than a full plan year. For example, for a person who turns 65 in July 2011, we observe age-64 diagnoses for only 6 months. This drives all risk scores down relative to official statistics or nationally representative data.<sup>50</sup> Nonetheless, these individual data have distinct advantages for considering longer-run equilibrium effects and for examining which reported conditions are driving the observed coding differences. We turn to each of these issues next.

### 6.3 Practice Style Spillovers in Equilibrium

One potential concern with the market-level analysis in Section 5 is that after some adjustment period, the increased presence of MA plans in a market could change local practice styles of FFS physicians in a way that causes coding under FFS to become more intense. For example, the same providers may treat both FFS and MA patients. For convenience these providers could choose to code both sets of patients similarly. As the share of the provider's patients covered by MA increased, her incentive to code *all her patients* according to MA coding incentives would also increase. Since we have examined a time of rapid expansion in MA in Section 5, this would cause our results to be overestimates of the long-run equilibrium differences if partial convergence in coding practices eventually occurred.

In contrast, the person fixed-effects strategy of this section estimates equilibrium coding differences in a way that fully captures longer-run effects after accounting for the presence of any local spillovers, because here, we do not rely on a changing local market structure. Instead, in the years leading up to our sample period for the Massachusetts analysis (2011-2012), MA plan presence in

---

<sup>50</sup>In addition, unlike our main analysis in Section 5, the person-fixed effects strategy would not be robust to selection on health trajectory. Also unlike our main analysis, which has national coverage at all ages of Medicare enrollment, we cannot exploit the same heterogeneity across plan types or estimate coding differences at ages other than 65.

Massachusetts was remarkably stable: MA penetration in Massachusetts *declined* by an insignificant 0.2 percentage points from 2008 to 2011. This contrasts with the national mean increase of 9.1 percentage points from 2006 to 2011, which identifies our estimates of effects over the 2007 to 2011 period in the national sample. Therefore, changes in FFS coding practices due to local practice style spillovers would have had time to equilibrate.

The person fixed-effects results based on the Massachusetts data and the age 65 threshold thus represent a test of whether our estimates from Section 5.1 are overestimates of the long-run equilibrium effects: Because the estimated coding differences are very similar under the two identification strategies, it provides evidence—consistent with beliefs and action by regulators—that FFS coding practices are not merely converging to MA coding practices over time.

## 6.4 Coding Margins

The Massachusetts data also allow us to examine in more detail which diagnosis codes drive the effect we document. We estimate difference-in-differences regressions similar to those presented in Table 8 but with indicators for specific chronic conditions that enter the risk-scoring algorithm as the dependent variables. For parsimony, we collapse the 70 condition categories used to generate risk scores into 24 mutually exclusive categories, following Pope et al. (2011).

In Figure 5, we plot the coefficients on  $\text{Post-65} \times \text{Selected MA}$  for the 24 condition groups, with each bar representing a separate regression with a different dependent variable. Conditions are ordered by the size of the coefficient estimate. The chronic condition categories displaying the largest differential prevalence in MA vs. FFS are infections, neoplasms (cancers and tumors), diabetes, and heart conditions. Although the results suggest substantial heterogeneity in upcoding across conditions, we detect no generalizable pattern.<sup>51</sup>

For two disease groups, diabetes and neoplasms, the regulator’s risk-scoring algorithm defines levels of severity of the disease that generate different payments—for example, “Diabetes without

---

<sup>51</sup>The results with respect to neoplasms (cancers and tumors) deserve special comment. Our analysis of the independently-collected SEER cancer data shows no evidence of cancer incidence responding to our identifying variation (Table 5). Only the cancer diagnoses as coded by the MA plans themselves show a significant effect. We also note that Appendix Figure A4 shows that the cancer effect is entirely driven by the lowest severity rating, which includes benign tumors and cancers in remission. When asked about this pattern, one MA physician explained that her note-taking software would pre-populate with past cancers and tumors and request her affirmation to be added to the patient’s record for the current plan year (regardless of the relevance of these tumors for the patient’s current health assessment). This type of action is unlikely to occur in FFS.

Complication” versus “Diabetes with Acute Complications.”<sup>52</sup> For diabetes, these levels are determined by the presence of diagnosed comorbidities and complications. For neoplasms, the levels of severity are determined by the site and malignancy, with the lowest severity category including non-malignant tumors. In Appendix Figure A4, we investigate whether the coding intensity effects we find for diabetes and neoplasms occur along the extensive margin of obtaining any code within the group, or along the intensive margin of increased severity within the group. The figure shows that there is little evidence of any effect on the intensive margin—that is, there is no shift toward higher severity codes within the disease categories. Instead, the largest effects seem to occur along the extensive margin. Additionally, the extensive margin effects are concentrated in the lowest level of severity within each category. This suggests that an important margin along which coding intensity varies may be the recording of a marginal disease, rather than increasing the reported severity within a disease category.

## 7 Discussion

### 7.1 Public Finance Impacts

Our coefficients can be used to calculate the implicit differential voucher available for enrolling in Medicare Advantage (MA) plans during our study period and today. As discussed in Section 2, this differential voucher implies a taxpayer-financed subsidy to more intensely-coded plans. The per-enrollee subsidy is equal to the benchmark rate multiplied by the difference in coding intensity ( $r_{i,MA} - r_{i,FFS}$ ), which we estimate to be 6.4% (0.064 in column 3, Table 2).

In 2014, the average annual benchmark was about \$10,000, implying that the additional costs to the government of MA enrollment would be around \$640 per person or \$10.5 billion total annually absent any correction.<sup>53</sup> In 2010, toward the end of our study period, CMS began deflating MA risk scores by 3.41% due to concerns about upcoding. In 2014, the deflation increased to 4.91%. 2014 risk scores thus remain on average 1.2% higher in MA compared to FFS after the deflation.<sup>54</sup> With 16 mil-

<sup>52</sup>In order of increasing severity, the cancer condition categories are Breast, Prostate, Colorectal, and Other Cancers and Tumors; Lymphatic, Head and Neck, Brain, and Other Major Cancers; Lung, Upper Digestive Tract, and Other Severe Cancers; and Metastatic Cancer and Acute Leukemia. In order of increasing severity, the diabetes condition categories are Diabetes without Complication; Diabetes with Ophthalmologic or Unspecified Manifestation; Diabetes with Acute Complications; Diabetes with Neurologic or Other Specified Manifestation; and Diabetes with Renal or Peripheral Circulatory Manifestation.

<sup>53</sup>The mean benchmark rate across counties for 2014 was \$10,020 for a plan with a 4-star (of 5) quality rating.

<sup>54</sup>The deflation is applied by multiplying the risk score by one minus the deflation factor:  $1.064 \times (1 - 0.0491) = 1.012$ .

lion Medicare beneficiaries choosing MA, even with the larger 2014 deflation this implies a remaining excess public cost of \$120 per MA enrollee or approximately \$2 billion in total in 2014.<sup>55</sup> While these 2014 excess costs are substantial, during our study period there was no coding adjustment (pre-2010) or a smaller adjustment (2010 and 2011), and the excess public cost of MA enrollment was much higher.

Our results in Tables 6 and 7 indicate that the policy of uniformly deflating MA risk scores fails to account for large coding differences across insurance contract types. Whereas PFFS MA plans differ in coding intensity by 5-6% relative to FFS, HMOs, which comprise the largest category of MA plans, inflate risk scores by 10%. And physician-owned plans inflate scores by 16%. This translates into an incremental implicit subsidy of about \$1000 and \$1,600 per MA enrollee annually to HMOs and physician-owned plans, respectively, before any coding deflation. The 2014 coding deflation is far too small for these plans, with overpayments still exceeding \$450 and \$1000, respectively, per MA enrollee per year.<sup>56</sup>

These costs of differential coding are substantial. Brown et al. (2014) study the uncompensated advantageous selection into MA after the implementation of risk adjustment in 2004. They estimate that the introduction of risk adjustment led to \$317 per enrollee in additional annual overpayments to MA, relative to the cost of insuring beneficiaries under FFS, because MA plans attracted relatively low-cost enrollees conditional on their risk scores. Our results, which would add to this \$317, 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 it does not consider the possibility of manipulable coding.

We have generally assumed that  $r_{i,MA} - r_{i,FFS}$  is constant across consumers or varies in a way that is uncorrelated with MA penetration,  $\theta^{MA}$ . If, contrary to our maintained assumption, individual-level heterogeneity in  $r_{i,MA} - r_{i,FFS}$  were correlated with  $\theta^{MA}$ , then our main results would capture coding differences only for the individuals marginal to our variation. In practice, these marginal types are likely to be close to the average MA enrollee. This is because the variation in  $\theta^{MA}$  we exploit in our empirical analysis covers most of the relevant range of MA penetration, as it arises

---

In 2015, the deflation is slated to increase again to 5.91%, close to our estimates.

<sup>55</sup>Based on September 2014 enrollment of 16,347,808 beneficiaries in MA, reported in Monthly SCC Enrollment Files provided by CMS.

<sup>56</sup>Coding differences net of the 2014 deflation are 4.6% for HMOs and 10.3% for physician-owned plans. Coding differences net of the smaller 2010-2013 deflation are 6.2% for HMOs and 12% for physician-owned plans.

from a period of massive expansion in the MA market segment. Therefore, even if the individual-specific coding differences are systematically different for beneficiaries who choose MA versus FFS, our estimates likely reflect the parameter necessary to calculate the excess public spending, which is the coding difference conditional on choosing MA, or  $E[r_{i,MA} - r_{i,FFS} \mid MA_i = 1]$ .

## 7.2 Coding Intensity and Consumer Choice

### 7.2.1 Theory

We next expand the model in Section 2 to trace out the implications of differential coding intensity for consumer choices. Our main goal in the paper is to identify coding intensity differences across insurers in Medicare and their impacts on public finances. Nonetheless, here we provide a more generalizable discussion of how differences in coding intensity across insurers distort consumer choices. This is particularly relevant for markets like the state and federal Health Insurance Exchanges, where the regulator implements risk adjustment by enforcing “budget-neutral” transfers from plans with below average risk scores to plans with above average risk scores. In this setting, there is no direct taxpayer cost to differential coding. Instead, coding differences affect social welfare by distorting consumer choices away from plans with low coding intensity and toward plans with high coding intensity.

Consider a general setting where consumers choose between two insurance contracts. Contract  $j$  consists of the characteristic set  $\{\gamma_j, \psi_j, \delta_j\}$ . As in 2,  $\gamma_j$  is a vector of attributes that affect risk scores but are not valued by consumers, and  $\psi_j$  is a vector of attributes that affect risk scores and are valued by consumers. We define  $\delta_j$  as a third vector of consumer-valued plan attributes that do not affect risk scores.  $\delta_j$  could include characteristics like network quality.

Define utility over plan  $j$  as  $v_{ij} = u_i(\psi_j, \delta_j)$ . Following the literature, we rule out income effects and assume price is additively separable from the utility derived from other plan characteristics.<sup>57</sup> With appropriate scaling of the  $u_i$  function, we can then describe the consumer choice problem between plans  $j$  and  $k$  as one of choosing  $j$  over  $k$  if and only if:

$$u_i(\psi_j, \delta_j) - p_j > u_i(\psi_k, \delta_k) - p_k. \quad (7)$$

<sup>57</sup>Many of the recent empirical studies of selection in insurance markets, including Einav, Finkelstein and Cullen (2010), Handel (2013), and Handel and Kolstad (2015) assume CARA preferences, a form which nests the assumption of no income effects.

where  $p$  gives the price that the consumer faces.

Next, consider the allocative efficiency condition, which requires that a consumer chooses plan  $j$  if and only if it represents the greatest utility net of the consumer's marginal cost in the plan  $(c_{ij}(\gamma_j, \psi_j, \delta_j))$ .<sup>58</sup> In Section 7.2.2, we discuss differential coding in the presence of multiple simultaneous market failures, but we narrow attention here to just the inefficiency introduced by differential diagnosis coding. To illustrate the impacts of differential coding even when risk adjustment works perfectly in counteracting selection inefficiencies, we make two simplifying assumptions. First, we assume marginal costs are additively separable in the plan and individual components:  $c_{ij}(\gamma_j, \psi_j, \delta_j) = \bar{c}_j(\gamma_j, \psi_j, \delta_j) + \hat{c}_i$ . This parameterization intentionally rules out phenomena like selection on moral hazard (Einav et al., 2013).<sup>59</sup> The efficiency condition is then:

$$u_i(\psi_j, \delta_j) - \bar{c}_j(\gamma_j, \psi_j, \delta_j) > u_i(\psi_k, \delta_k) - \bar{c}_k(\gamma_k, \psi_k, \delta_k), \quad (8)$$

where  $\hat{c}_i$  has canceled from both sides of the inequality, and the remaining  $\bar{c}$  is a function only of plan characteristics. This expression makes it clear that the first-best price differential that sorts consumers efficiently across plans is equal to the incremental marginal cost,  $\bar{c}_j(\gamma_j, \psi_j, \delta_j) - \bar{c}_k(\gamma_k, \psi_k, \delta_k)$ , following the standard treatment (e.g. Cutler and Reber, 1998; Einav, Finkelstein and Cullen, 2010). Note that this first-best price can be affected by any of the three categories of plan attributes.

Now, we turn to the question of how the equilibrium price may differ from the first-best price when coding differs across plans. Recall that using our notation from Section 2 we can write an individual's risk score as

$$r_{ij} = \hat{r}_i + \alpha(\gamma_j, \psi_j), \quad (9)$$

where  $\alpha(\gamma, \psi)$  describes a plan-specific coding factor that is a function of plan characteristics. Also recall that for each enrollee a plan receives a subsidy equal to a benchmark payment,  $\phi$ , multiplied by the enrollee's risk score,  $r_{ij}$ . We assume that in the absence of differential coding, the equilibrium price differential net of risk adjustment subsidies are equal to first best prices and would thus sort con-

<sup>58</sup>The (standard) welfare function is  $W = \sum_{i=1}^N \{(u_i(\psi_j, \delta_j) - c_{ij})1[\text{Choose } j] + (u_i(\psi_k, \delta_k) - c_{ik})1[\text{Choose } k]\}$ .

<sup>59</sup>This assumption also has the effect of generating identical first-best prices for every consumer, allowing us to abstract from forms of selection that cause no single price to sort consumers efficiently across plans as in Bundorf, Levin and Mahoney (2012) and Geruso (2012). Such forms of selection add complexity to describing the choice problem without providing additional insights into the consequences of coding differences for consumer choices.

sumers efficiently across plans. In other words, we assume that  $p_j - p_k = \bar{c}_j(\gamma_j, \psi_j, \delta_j) - \bar{c}_k(\gamma_k, \psi_k, \delta_k)$  whenever  $\alpha_j = \alpha_k$ .<sup>60</sup> This assumption asserts that risk adjustment succeeds in flattening the insurer's perceived cost curve in a competitive equilibrium setting (by setting  $\hat{c}_i = \phi \hat{r}_i$ ), consistent with the regulatory intention of risk adjustment.<sup>61</sup> Under these assumptions, prices are equal to the plan component of costs minus the plan component of the risk adjustment subsidy,  $p_j = c_{ij}(\gamma_j, \psi_j, \delta_j) - \phi \alpha(\gamma_j, \psi_j)$ , allowing us to re-write Eq. 7 as

$$u_i(\psi_j, \delta_j) - \bar{c}_j(\gamma_j, \psi_j, \delta_j) + \phi \alpha(\gamma_j, \psi_j) > u_i(\psi_k, \delta_k) - \bar{c}_k(\gamma_k, \psi_k, \delta_k) + \phi \alpha(\gamma_k, \psi_k). \quad (10)$$

The expression for choice in Eq. 10 differs from the efficiency condition in Eq. 8 by the term  $\phi \alpha(\gamma, \psi)$ , which captures the potential price distortion caused by the portion of the risk adjustment payment that depends on plan, rather than person, characteristics. To illustrate the source of the inefficiency, consider the consumer choice problem under three scenarios: (a) Plans differ in characteristics that do not affect risk scores ( $\delta_j \neq \delta_k$ ); (b) plans differ in characteristics that affect risk scores, but not consumer utility ( $\gamma_j \neq \gamma_k$ ); and (c) plans differ in characteristics that simultaneously affect risk scores and consumer utility ( $\psi_j \neq \psi_k$ ). Under these scenarios, the consumer choice problems from Eq. 10 are to choose  $j$  when:

$$u_i(\delta_j) - \bar{c}_j(\delta_j) + \phi \alpha > u_i(\delta_k) - \bar{c}_k(\delta_k) + \phi \alpha \quad \text{no coding differences} \quad (11a)$$

$$\mu_i - \bar{c}_j(\gamma_j) + \phi \alpha(\gamma_j) > \mu_i - \bar{c}_k(\gamma_k) + \phi \alpha(\gamma_k) \quad \text{valueless, costly coding differences} \quad (11b)$$

$$u_i(\psi_j) - \bar{c}_j(\psi_j) + \phi \alpha(\psi_j) > u_i(\psi_k) - \bar{c}_k(\psi_k) + \phi \alpha(\psi_k), \quad \text{valued, costly coding differences} \quad (11c)$$

where notation is suppressed wherever characteristics are identical across plans. Only for the case in which coding is identical across plans, (a), is the first-best price achieved, causing the choice problem to match the efficiency condition in Eq. 8.

In (b) and (c), choices are distorted away from the efficiency condition by the  $\phi \alpha(\gamma, \psi)$  terms. In

<sup>60</sup>This efficient price differential could be supported in a competitive equilibrium. The zero profit condition requires that prices are set equal to average costs net of risk adjustment. For example, let  $p_j = E[c_{ij}(\gamma_j, \psi_j, \delta_j) - \phi R_{ij} \mid \text{Choose } j] = E[\bar{c}_j(\gamma_j, \psi_j, \delta_j) + \hat{c}_i - \phi \hat{r}_i \mid \text{Choose } j]$ . If risk adjustment succeeds in flattening the firm's cost curve exactly as policymakers intend (and with no upcoding), then the person-specific components of plan costs are exactly compensated by the person-specific component of the risk adjustment payment ( $\hat{c}_i = \phi \hat{r}_i$ ), and the last terms cancel. In that case  $p_j = \bar{c}_j(\gamma_j, \psi_j, \delta_j)$ , which satisfies zero profits.

<sup>61</sup>This assumption is not necessary to derive the results in this section, but allows us significant economy of notation



(b), differential coding affects the insurer's costs and risk-adjusted payments but does not affect consumer valuation. Therefore, consumers value both plans equally and simply choose the plan with the lowest net price,  $\bar{c} - \phi\alpha$ , even if that plan represents greater social cost. In (c), the plan characteristics that consumers value simultaneously affect coding. This would be the case, for example, if lower copays impacted utilization (which is valued by consumers) and also affected the probability that a diagnosis was recorded.

Note that comparing across the three scenarios, the first-best prices, and thus the optimal allocations of consumers to plans, defined by  $u_i(\cdot) - \bar{c}_j(\cdot)$ , would differ. This is because the coding activity changes the plans' marginal costs as well as the consumers' valuations. Nevertheless, in (b) and (c) the differential subsidy ( $\alpha(\gamma_j, \psi_j) - \alpha(\gamma_k, \psi_k)$ ) distorts away from whatever is the relevant optimum. This is true for case (c) even though the coding differences there are due entirely to plan attributes that consumers value.

The intuition here is straightforward: Like any other market setting, differentially subsidizing a particular consumer choice reduces total surplus unless the subsidy counteracts another market failure. In principle, the distortion of consumer choices toward more intensely coded plans could be efficient (in a second-best sense) if it counteracted other distortions operating simultaneously that caused under-subscription of consumers to those same intensely coded plans. We discuss this possibility in the next section for some commonly considered market failures in Medicare, though the point here is that there is no *a priori* reason to think subsidizing high coding intensity is even second best. Finally, the model shows that with respect to consumer choice distortions, an important parameter is the complete differential voucher,  $\phi r_{ij} - \phi r_{ik} = \phi(\alpha(\gamma_j, \psi_j) - \alpha(\gamma_k, \psi_k))$ , which we calculated in Section 7.1.

### 7.2.2 Choice Distortions Implied by Estimates

To understand the extent to which the differential subsidy we estimate above affects consumer choices in our empirical setting, we use estimates of MA demand elasticities from the well-developed prior literature. We consider a hypothetical policy that removes the subsidy currently given for intense coding in MA. Such a policy would have the effect of reducing the plan payment rate.<sup>62</sup> Formally, we

---

<sup>62</sup>This hypothetical policy removes only the overpayments due to coding. It does not remove, for example, the overpayments of MA relative to FFS that were intentionally generated by Congress, for example to promote the MA program in rural areas (Achman and Gold, 2002).

consider uniformly deflating the risk-adjusted payment from the regulator to plan  $j$  to compensate for the mean MA-FFS coding difference ( $\bar{R}_j = \frac{\phi \cdot \bar{r}_j}{1.064}$ ).<sup>63</sup> In the conclusion, we discuss more nuanced refinements of current risk adjustment policy, with the potential to reduce within-MA coding differences.

Most previous studies of MA demand have estimated plan price semi-elasticities  $\left(\epsilon_P \equiv \frac{\partial \theta}{\partial P} \cdot \frac{1}{\theta}\right)$ , where  $P$  equals the consumer premium and, as above,  $\theta$  denotes MA market segment enrollment as a share of total Medicare enrollment. Table 9 reports these demand elasticities from the literature. The parameter needed for generating the relevant counterfactuals is the elasticity of demand for the MA market segment (as a whole) with respect to payments to MA plans. We convert plan price semi-elasticities to plan payment semi-elasticities  $\left(\epsilon_\phi \equiv \frac{\partial \theta}{\partial \phi} \cdot \frac{1}{\theta}\right)$  by multiplying by the pass-through rate,  $\rho = -\frac{\partial P}{\partial \phi}$ , which expresses how the marginal government payment to MA plans translates into lower premiums or higher rebates for consumers. Theory predicts that under perfect competition and assuming no selection on enrollee net cost, this parameter would equal 1, as competition forced premiums down dollar-for-dollar with the increased subsidy. A wide range of studies, including Song, Landrum and Chernew (2013), Cabral, Geruso and Mahoney (2014), and Curto et al. (2014), find pass-through rates in MA of about 50%.<sup>64</sup> We therefore assume 50% pass-through, apply a 6.4% upcoding overpayment (Table 2, col 3), and calculate counterfactual enrollment changes assuming this overpayment was removed.

The back-of-the-envelope calculations in the third column of Table 9 show that modifying risk adjustment payments in this way would have large negative impacts on MA enrollment for the range of elasticities estimated in the literature. Using the smallest demand elasticity (Cabral, Geruso and Mahoney, 2014) yields a 17% decline in the size of the MA market, while the most elastic estimate (Dunn, 2010) implies a 33% decline in the size of MA under the counterfactual.

These estimates of reductions in MA enrollment apply to a complete elimination of the subsidy relative to a counterfactual where CMS does not adjust risk scores to account for upcoding. As discussed above, at the end of our sample period CMS began deflating risk scores by 3.41%. In the last column of Table 9 we show the enrollment effects of increasing the CMS risk score deflation from 3.41% to the 6.4% coding difference we estimate. The estimates suggest that the additional deflation

<sup>63</sup>See Eq. 2.

<sup>64</sup>A notable exception is Duggan, Starc and Vabson (2014), which finds a point estimate of zero pass-through, but with confidence intervals that include 50%.

would result in a decline in MA enrollment of between 8% and 15% (a decline of 2 and 4 percentage points, respectively, based on 2010 enrollment).<sup>65</sup>

If, in addition to an all-MA deflation, coding were differentially deflated across plan types to properly account for coding intensity differences within MA, it is likely that the sorting of consumers across plans within the MA segment would be dramatically affected. Provider-owned plans currently capture the largest coding subsidies. Leveling payments would imply reducing payments to these plans by an additional 9.5% beyond the all-MA payment reductions ( $= 15.6 - 6.1$ , from Table 7, column 3). Curto et al. (2014) estimate demand response at the plan level, rather than MA market segment level, and find own-price elasticities around -1, implying this kind of reshuffling within the MA market could be substantial.

A non-obvious implication of our modeling above is that stronger competition may actually exacerbate the distortionary consequences of differential coding. If competition is strong, then the pass-through rate will be higher, resulting in higher plan payment semi-elasticities and a stronger relationship between the coding subsidy and enrollment. While higher pass-through increases consumer surplus at the cost of producer surplus, this higher pass-through could also decrease net efficiency by more significantly distorting choices. On the other hand, if competition is weaker, insurers will pass-through a smaller fraction of the overpayment, resulting in a smaller distortion to consumers' choices between FFS and MA. In summary, increased competition tilts the incidence of the transfer/subsidy toward consumers but can simultaneously worsen the choice distortion.

While our estimates imply that the effects of the coding subsidy on MA enrollment are substantial, the welfare consequences remain unclear. Thus far, we have considered the differential coding distortion in isolation. In a setting with multiple simultaneous market failures, the question of efficiency depends on whether the coding distortion counteracts some other opposing market failure. For example, if MA insurers have market power, then markups may raise prices above costs. In this case, the differential voucher due to the unrelated phenomenon of upcoding could be welfare-improving. Of course, this offsetting relationship between "under-enrollment" in MA due to imperfect competition and "over-enrollment" in MA due to the coding subsidy would be purely coincidental. There is no reason *a priori* to expect that the coding subsidy specifically counteracts MA

---

<sup>65</sup>Because these implied effects are large, we note that the only parameter from the present study entering into the calculations in Table 9 is the 6.4% mean coding difference, estimated robustly via two distinct research designs and consistent with regulators' policy actions and long held beliefs about coding.

insurer market power or any of the other important market failures documented in the context of Medicare Advantage, such as inertia (Sinaiko, Afendulis and Frank, 2013), spillovers between MA and FFS (Baicker, Chernew and Robbins, 2013), or the choice salience of non-cash rebates (Stockley et al., 2014).

## 8 Conclusion

The manipulability of the risk adjustment system via diagnosis coding is an issue of significant practical importance, given the large and growing role of risk adjustment in regulated insurance markets. Our results demonstrate wide scope for upcoding in Medicare Advantage, one of the largest risk-adjusted health insurance markets in the US, relative to the fee-for-service option. The estimates imply significant overpayments to private insurers at a cost to the taxpayer, as well as distortions to consumer choices. We also find strong evidence that coding intensity is increasing in a plan's level of insurer-provider integration.

The phenomenon we explore here is relevant in other risk-adjusted markets, such as the ACA Exchanges. In the Exchanges, private insurers do not compete with a public option. 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>66</sup> Nonetheless, even budget-neutral transfers generate the same kind of *differential* implicit subsidies discussed in the context of Medicare. Our result that MA plans with higher levels of insurer-provider integration display higher coding intensity suggests that there will be asymmetric coding among Exchange plans as well, with HMOs and other integrated insurer-provider organizations coding most intensely. This will distort the choices of Exchange enrollees toward the more integrated plans.

Nonetheless, 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. Our study offers some insight into potential improvements in risk adjustment mechanism design: From the per-

---

<sup>66</sup>A caveat is that in markets such as the Exchanges where the government pays subsidies based on the premiums set by insurers, there could 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 premium—higher subsidies and additional government spending.

spective 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 health plan benefit design, risk adjustment must focus on generating insurer incentives rather than predicting expected costs. Applied to our findings, this insight suggests that the (second-best) optimal payment policy may include risk adjustment, but with coefficients on risk adjusters that account for the susceptibility of each code to differential coding (of which we provide evidence in [Figure 5](#)).

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 includes coding distortions. In practice, because the upcoding susceptibility of risk adjusters (especially new risk adjusters) may be difficult to observe, fewer and coarser diagnosis groups might be a preferable feasible alternative to current risk adjustment systems. Another simple change that could reduce the impact of coding heterogeneity across insurers would be to expand the look-back period over which risk scores are calculated. Insurers increase risk scores in part by attempting to reassign prior-year diagnoses—for example, by reviewing charts automatically, prompting providers with past diagnoses, and/or by proactively sending providers to the homes of patients with particular medical histories. A longer look-back period would reduce heterogeneity in coding along this margin, particularly in FFS Medicare (in which providers have little incentive to retain prior diagnoses) versus MA.

Even with significant reform, it may not be possible to achieve perfect parity of coding intensity across the Medicare Advantage and FFS market segments or within MA, among its competing plans. In that case, any benefits of the MA program in terms of generating consumer surplus or creating cost-saving externalities within local healthcare markets (as in [Baicker, Chernew and Robbins, 2013](#)) should be weighed against the additional taxpayer costs and consumer choice distortions generated by a regulatory system in which the parameters determining insurer payment are squishy.

## References

- Achman, Lori, and Marsha Gold.** 2002. Medicare+ Choice 1999-2001: An Analysis of Managed Care Plan Withdrawals and Trends in Benefits and Premiums. Commonwealth Fund.
- 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.
- Bundorf, M. Kate, Jonathan Levin, and Neale Mahoney.** 2012. "Pricing and Welfare in Health Plan Choice." American Economic Review, 102(7): 3214–48.
- Cabral, Marika, Michael Geruso, and Neale Mahoney.** 2014. "Does Privatized Health Insurance Benefit Patients or Producers? Evidence from Medicare Advantage." NBER Working Paper.
- Carey, Colleen.** 2014. "Government Payments and Insurer Benefit Design in Medicare Part D."
- Chetty, Raj, John N. Friedman, 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.
- Clemens, Jeffrey, and Joshua D Gottlieb.** 2014. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" The American economic review, 104(4): 1320.
- 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.
- Curto, Vilsa, Liran Einav, Jonathan Levin, and Jay Bhattacharya.** 2014. "Can Health Insurance Competition Work? Evidence from Medicare Advantage." National Bureau of Economic Research Working Paper 20818.
- Cutler, David M., and Sarah J. Reber.** 1998. "Paying for Health Insurance: The Trade-Off between Competition and Adverse Selection." The Quarterly Journal of Economics, 113(2): 433–466.
- Dafny, Leemore S.** 2005. "How Do Hospitals Respond to Price Changes?" American Economic Review, 1525–1547.
- Decarolis, Francesco.** 2015. "Medicare Part D: Are Insurers Gaming the Low Income Subsidy Design?" American Economic Review, 105(4): 1547–80.
- Duggan, Mark, Amanda Starc, and Boris Vabson.** 2014. "Who Benefits When the Government Pays More? Pass-Through in the Medicare Advantage Program." NBER Working Paper.
- Dunn, Abe.** 2010. "The Value of Coverage in the Medicare Advantage Insurance Market." Journal of Health Economics, 29(6): 839–855.

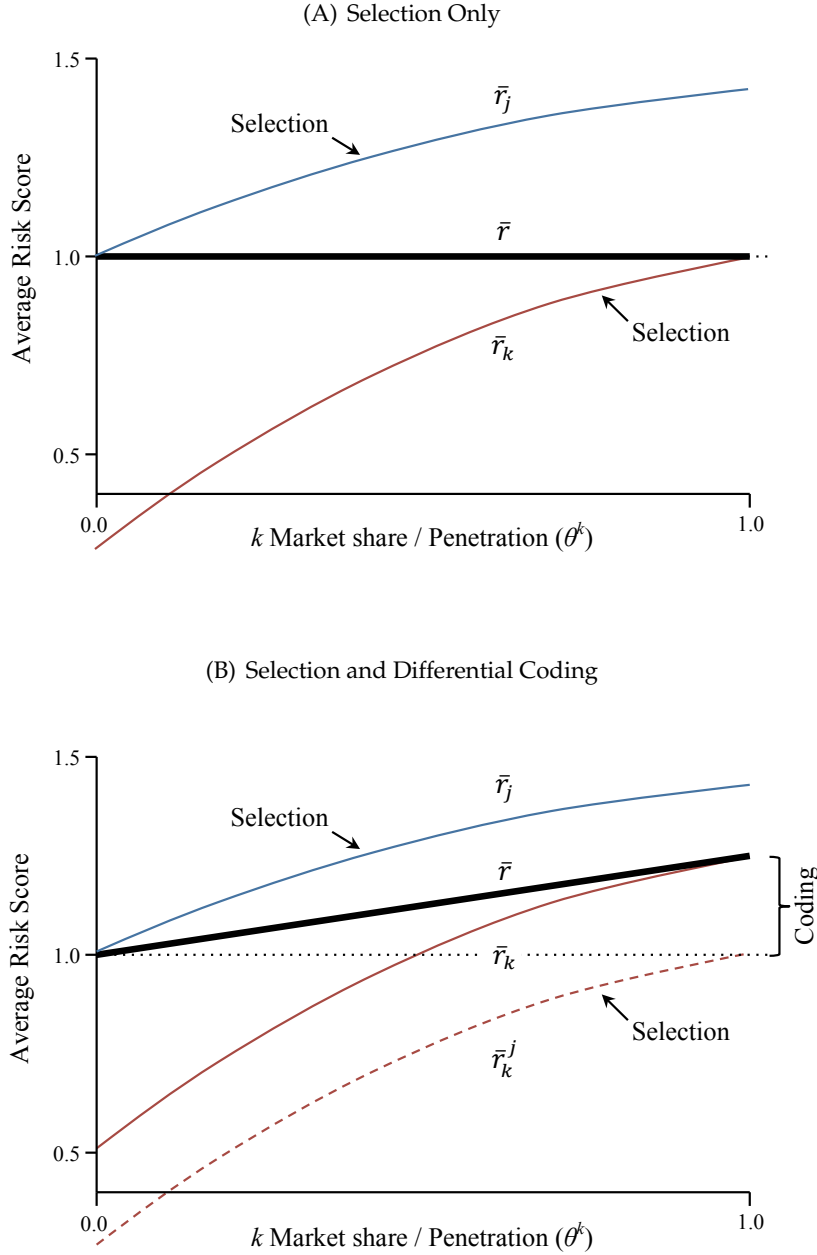
- Einav, Liran, Amy Finkelstein, and Mark R. Cullen.** 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices." The Quarterly Journal of Economics, 125(3): 877–921.
- Einav, Liran, Amy Finkelstein, Raymond Kluender, and Paul Schrimpf.** 2015. "Beyond statistics: the economic content of risk scores." MIT.
- Einav, Liran, Amy Finkelstein, Stephen P. Ryan, Paul Schrimpf, and Mark R. Cullen.** 2013. "Selection on Moral Hazard in Health Insurance." American Economic Review, 103(1): 178–219.
- Einav, Liran, and Jonathan Levin.** 2014. "Managed Competition in Health Insurance."
- Fisher, Elliott S, Stephen M Shortell, Sara A Kreindler, Aricca D Van Citters, and Bridget K Larson.** 2012. "A framework for evaluating the formation, implementation, and performance of accountable care organizations." Health Affairs, 31(11): 2368–2378.
- Frakt, Austin B, and Rick Mayes.** 2012. "Beyond capitation: how new payment experiments seek to find the 'sweet spot' in amount of risk providers and payers bear." Health Affairs, 31(9): 1951–1958.
- Frakt, Austin B, Steven D Pizer, and Roger Feldman.** 2013. "Plan–Provider Integration, Premiums, and Quality in the Medicare Advantage Market." Health Services Research, 48(6pt1): 1996–2013.
- Frandsen, Brigham, and James B Rebitzer.** 2014. "Structuring Incentives within Accountable Care Organizations." Journal of Law, Economics, and Organization, ewu010.
- 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.
- Gaynor, Martin, James B Rebitzer, and Lowell J Taylor.** 2004. "Physician incentives in health maintenance organizations." Journal of Political Economy, 112(4): 915–931.
- Gaynor, M., K. Ho, and R. Town.** 2015. "The Industrial Organization of Health Care Markets." Journal of Economic Literature, 53(2).
- Geruso, Michael.** 2012. "Selection in Employer Health Plans: Homogeneous Prices and Heterogeneous Preferences."
- Geruso, Michael, and Thomas McGuire.** 2015. "Tradeoffs in the Design of Health Plan Payment Systems: Fit, Power and Balance." NBER Working Paper.
- 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>.



- Hackmann, Martin B, Jonathan T Kolstad, and Amanda Ellen Kowalski.** 2015. "Adverse Selection and an Individual Mandate: When Theory Meets Practice." The American Economic Review, 105(3): 1030–1066.
- Handel, Benjamin R.** 2013. "Adverse selection and inertia in health insurance markets: When nudging hurts." The American Economic Review, 103(7): 2643–2682.
- Handel, Benjamin R, and Jonathan T Kolstad.** 2015. "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare." American Economic Review, (forthcoming).
- Ho, Kate, Joe Hogan, and Fiona Scott Morton.** 2014. "The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program."
- Kolstad, Jonathan T.** 2013. "Information and Quality When Motivation Is Intrinsic: Evidence from Surgeon Report Cards." American Economic Review, 103(7): 2875–2910.
- Kronick, Richard, and W. Pete Welch.** 2014. "Measuring Coding Intensity in the Medicare Advantage Program." Medicare and Medicaid Research Review, 4(2): E1–E19.
- Newhouse, Joseph P, Mary Price, Jie Huang, J Michael McWilliams, and John Hsu.** 2012. "Steps to reduce favorable risk selection in Medicare advantage largely succeeded, boding well for health insurance exchanges." Health Affairs, 31(12): 2618–2628.
- Pope, Gregory C, John Kautter, Melvin J Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart.** 2011. "Evaluation of the CMS-HCC Risk Adjustment Model." CMS Division of Risk Adjustment and Payment Policy Final Report March.
- Pope, Gregory C, John Kautter, Randall P Ellis, Arlene S Ash, John Z Ayanian, Melvin J Ingber, Jesse M Levy, and John Robst.** 2004. "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model." Health Care Financing Review, 25(4): 119.
- 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.
- Sacarny, Adam.** 2014. "Technological Diffusion Across Hospitals: The Case of a Revenue-Generating Practice." MIT.
- Schulte, Fred.** 2014. "Judgment calls on billing make upcoding prosecutions rare." The Center for Public Integrity, May 19, 2014. Available: <http://www.publicintegrity.org/2012/09/15/10835/judgment-calls-billing-make-upcoding-prosecutions-rare> [Last accessed: 15 April 2015].
- Schulte, Fred.** 2015. "More Whistleblowers Say Health Plans Are Gouging Medicare." National Public Radio, May 19, 2014. Available: <http://www.npr.org/blogs/health/2015/04/23/401522021/more-whistleblowers-say-health-plans-are-gouging-medicare> [Last accessed: 15 April 2015].
- 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?" INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 50(3): 202–215.

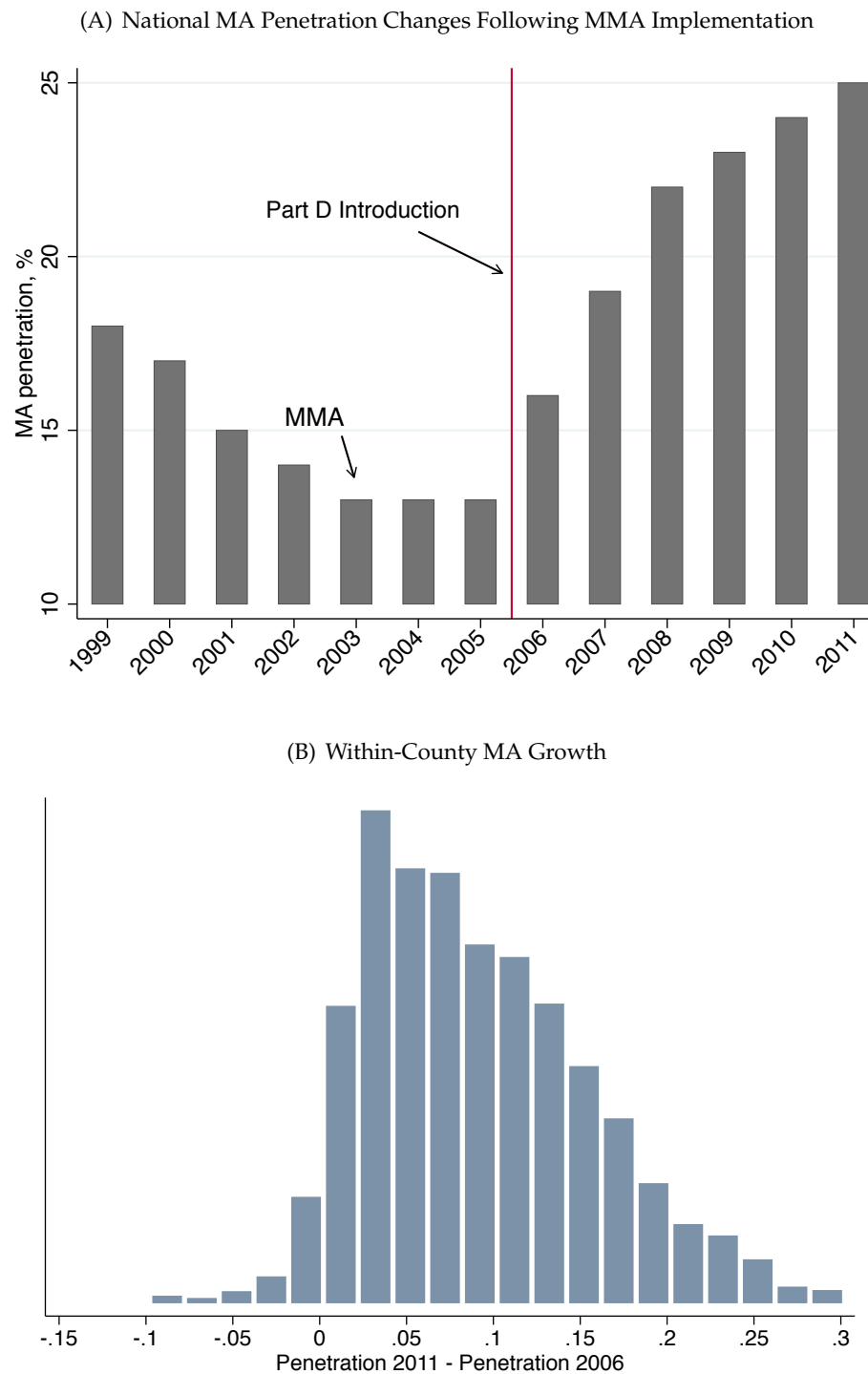
- Song, Yunjie, Jonathan Skinner, Julie Bynum, Jason Sutherland, John E Wennberg, and Elliott S Fisher.** 2010. "Regional variations in diagnostic practices." New England Journal of Medicine, 363(1): 45–53.
- Song, Zirui, Mary Beth Landrum, and Michael E Chernew.** 2013. "Competitive bidding in Medicare Advantage: Effect of benchmark changes on plan bids." Journal of health economics, 32(6): 1301–1312.
- Song, Zurui, and Jacob Wallace.** 2015. "The Impact of Turning 65 on Spending, Volume and Price."
- Stockley, Karen, Thomas McGuire, Christopher Afendulis, and Michael E. Chernew.** 2014. "Premium Transparency in the Medicare Advantage Market: Implications for Premiums, Benefits, and Efficiency." NBER Working Paper.

**Figure 1: Identifying Coding Differences in Selection Markets**



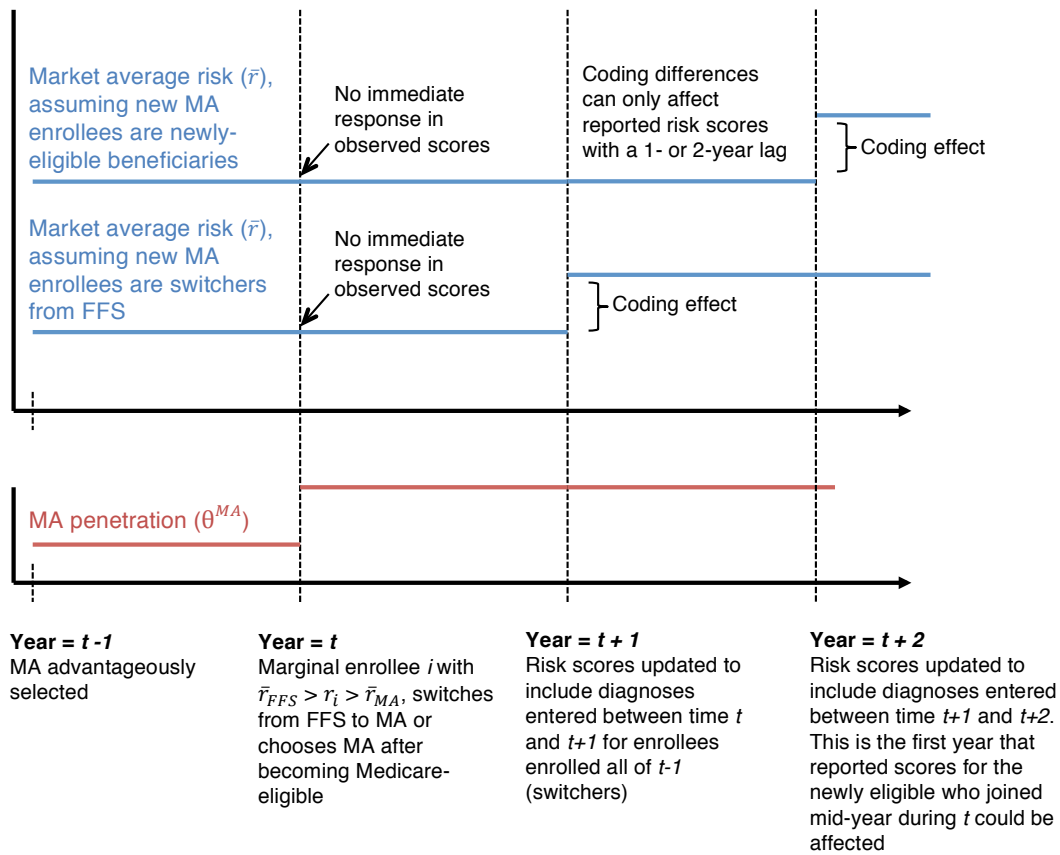
**Note:** The figure illustrates how to separate coding intensity differences from selection when true underlying health risk is unobservable. The horizontal axis measures the market share of plan (or market segment)  $k$ ,  $\theta^k$ . The vertical axis measures the average risk score: Average risk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

**Figure 2: Identification: Within-County Growth in Medicare Advantage Penetration**



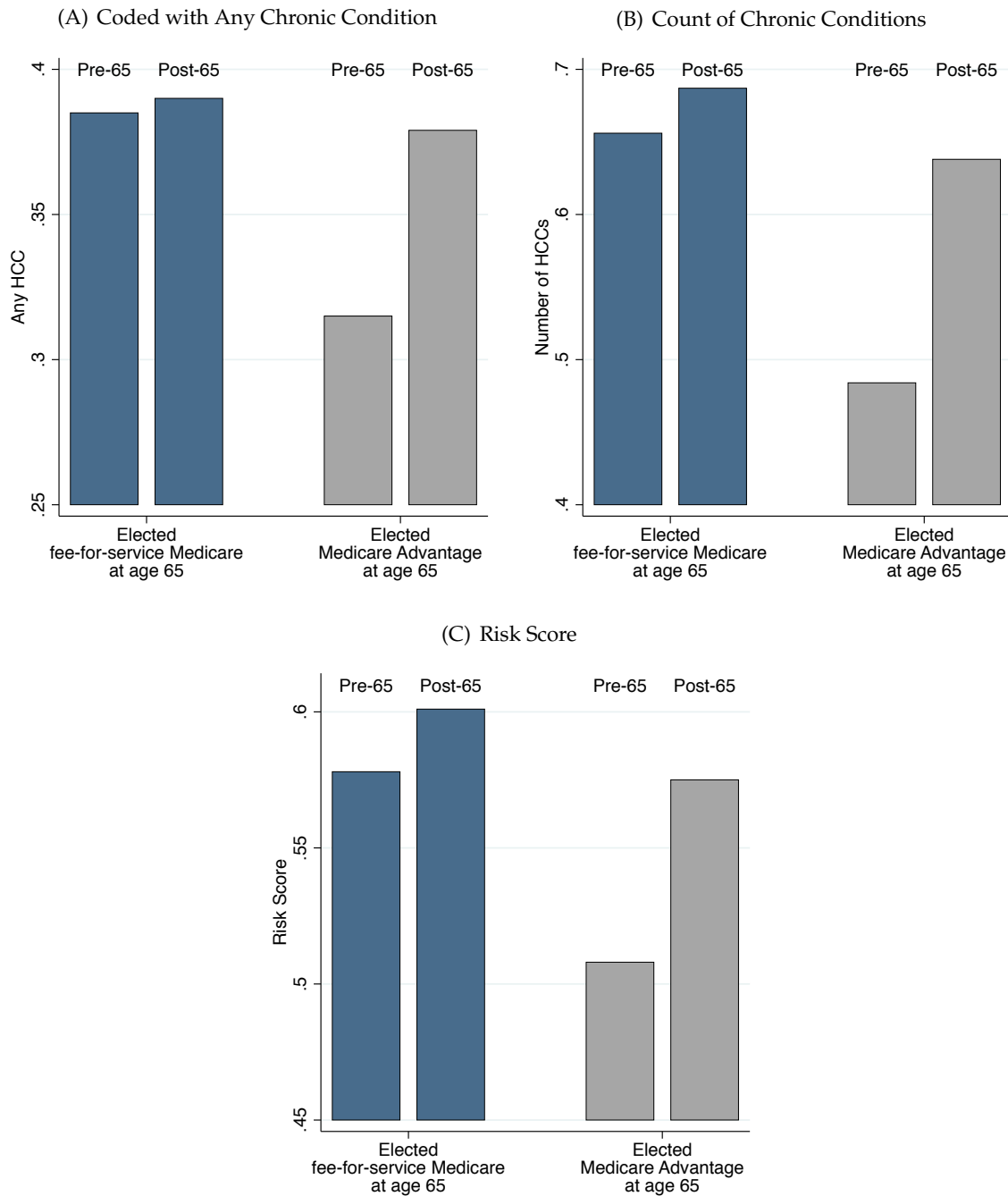
**Note:** The top panel displays national trends in MA penetration, where the unit of observation is the Medicare beneficiary. *Source:* Kaiser Family Foundation, 2013. The bottom panel displays a histogram of within-county changes in penetration from 2006 to 2011 in the main estimation sample. The unit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

**Figure 3: Identification: Coding Effects Plausibly Observed Only with a Lag in Medicare**



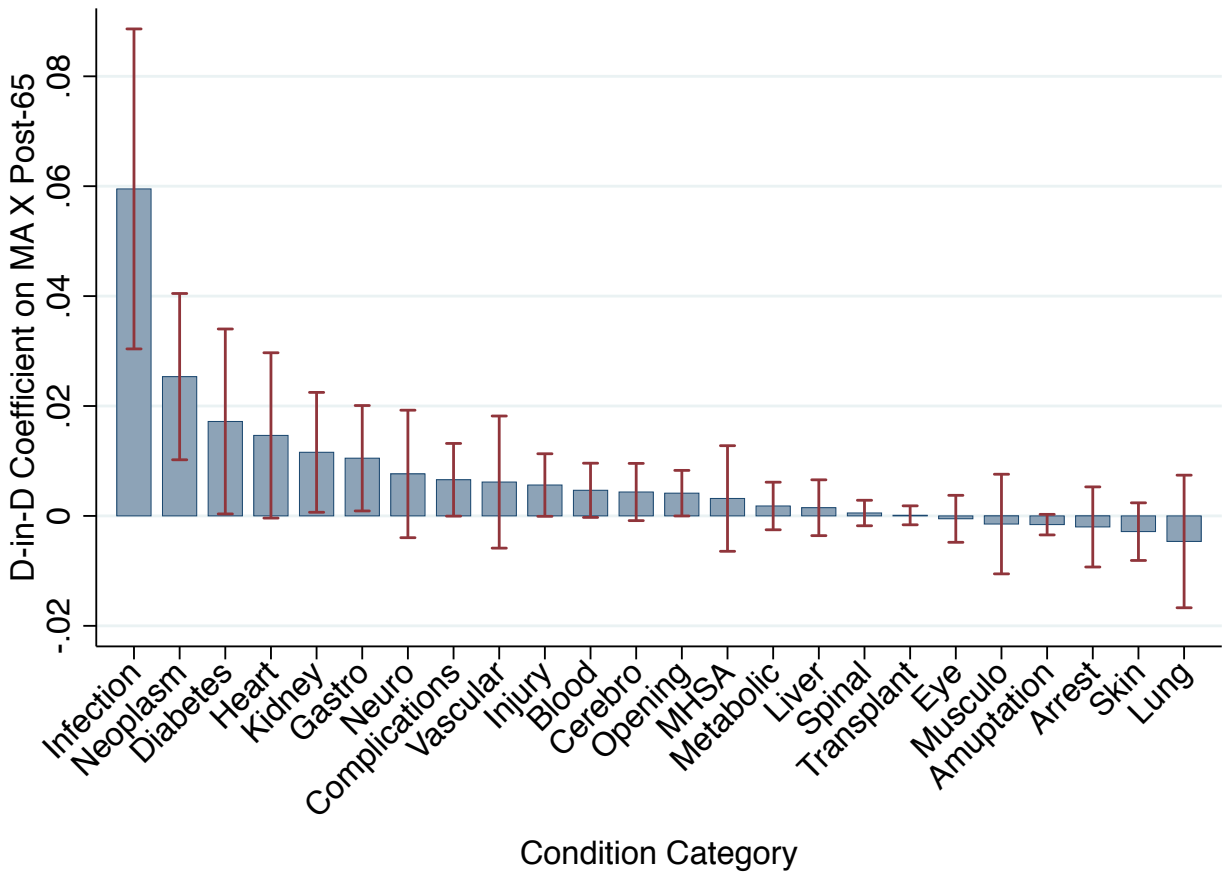
**Note:** The diagram highlights the timing of changes in market-level average risk scores ( $\bar{r}$ ) in response to a change in MA penetration ( $\theta^{MA}$ ). For the first year in either the MA or FFS market segment, switchers carry forward risk scores based on their diagnoses from the previous year in the other segment. For the newly eligible beneficiaries (those turning 65), demographic risk scores are assigned until there is a full calendar year of enrollment and diagnosis information. Therefore, effects on risk scores should not be apparent until year  $t+1$  when the net penetration change is due to switchers and should not be apparent until year  $t+2$  when the net penetration change is due to new Medicare enrollees.

**Figure 4:** Alternative Identification: Coding Across the Age 65 Threshold, MA vs. FFS



**Note:** The figure shows difference-in-differences summary statistics comparing the change in an individual's risk score and diagnosed conditions at age 65 among consumers entering MA and consumers entering FFS. Blue bars indicate individuals who chose to enroll in FFS at age 65, and gray bars indicate individuals who chose to enroll in MA at age 65. Data are from the Massachusetts All-Payer Claims Database, as described in the text. Prior to age 65, all consumers in the sample are continuously enrolled in commercial/employer plans. We describe the construction of these statistics in more detail in Appendix A.6.

**Figure 5:** Coding Differences by Condition Across the Age 65 Threshold, MA vs. FFS



**Note:** The figure shows coefficients from difference-in-differences regressions in which the dependent variable is an indicator for being diagnosed with a specific condition. These regressions compare pre and post diagnoses among consumers aging into Medicare and selecting MA versus FFS. Each bar represents the coefficient on  $\text{Post-65} \times \text{Selected MA}$  from a separate OLS regression. The specification parallels the regressions in Table 8. The set of conditions included are those that enter the risk score calculation. Red bars indicate the 95% confidence intervals. Data are from the Massachusetts All-Payer Claims Database, as described in the text. Prior to age 65, all consumers in the sample are continuously enrolled in commercial/employer plans. We describe the construction of these statistics in more detail in Appendix A.6.



**Table 1: 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%	0.0%	3128
MA-Part D penetration	5.3%	8.0%	13.1%	10.8%	3128
MA non-Part D penetration	1.8%	3.0%	3.0%	4.0%	3128
Market risk score	1.000	0.079	1.000	0.085	3128
Risk score in TM	1.007	0.082	1.003	0.084	3128
Risk score in MA	0.898	0.171	0.980	0.147	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:** The table shows county-level summary statistics for the first and last year of the main analysis sample. The sample consists of 3,128 counties for which we have a balanced panel of data on Medicare Advantage 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 Medicare months in the county  $\times$  year. The market risk score is averaged over all Medicare beneficiaries in the county and normed to 1.00 nationally in each year.

**Table 2: Main Results: Impacts of MA Expansion on Market-Level Reported Risk**

	Dependent Variable: County-Level Average Risk Score		
	(1)	(2)	(3)
MA Penetration $t$ (placebo)	0.007 (0.015)	0.001 (0.019)	0.001 (0.019)
MA Penetration $t-1$	0.069** (0.011)	0.067** (0.012)	0.064** (0.011)
Main Effects			
County FE	X	X	X
Year FE	X	X	X
Additional Controls			
State X Year Trend		X	X
County X Year Demographics			X
Mean of Dep. Var.	1.00	1.00	1.00
Observations	15,640	15,640	15,640

**Note:** The table shows coefficients from difference-in-differences regressions in which the dependent variable is the average risk score in the market (county). Effects of contemporaneous ( $t$ ) and lagged ( $t - 1$ ) Medicare Advantage (MA) penetration are displayed. Because MA risk scores are calculated using diagnosis data from the prior plan year, changes in MA enrollment can plausibly affect reported risk scores via differential coding only with a lag. Thus, contemporaneous penetration serves as a placebo that allows for a test for pre-trends within the county. Observations are county  $\times$  years. All specifications include county and year fixed effects. Column 2 additionally controls for state indicators interacted with a linear time trend. Column 3 additionally controls for the demographic makeup of the county  $\times$  year by including 18 indicator variables capturing the fraction of the population in 5-year age bins from 0 to 85 and  $>85$ . Standard errors in parentheses are clustered at the county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table 3: Placebo Tests and Lagged Effects**

	Dependent Variable: County-Level Average Risk Score			
	Placebo Tests			Lagged Effects
	available panel years: 2007-2011	2007-2010	2007-2009	2008-2011
	(1)	(2)	(3)	(4)
MA Penetration $t+2$ (placebo)			0.044+ (0.023)	
MA Penetration $t+1$ (placebo)		0.017 (0.025)	0.032 (0.056)	
MA Penetration $t$ (placebo)	0.001 (0.019)	-0.021 (0.028)	-0.064 (0.071)	0.006 (0.017)
MA Penetration $t-1$	0.064** (0.011)	0.076** (0.018)	0.084** (0.022)	0.041** (0.015)
MA Penetration $t-2$				0.046* (0.022)
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 X Year Demographics	X	X	X	X
Observations	15,640	12,512	9,384	12,512

**Note:** The table shows coefficients from difference-in-differences regressions in which the dependent variable is the average risk score in the market (county). Effects of future ( $t+2$ ,  $t+1$ ), contemporaneous ( $t$ ), and lagged ( $t-1$ ,  $t-2$ ) Medicare Advantage (MA) penetration are displayed. Because MA risk scores are calculated using diagnosis data from the prior plan year, changes in MA enrollment can plausibly affect reported risk scores via differential coding only with a lag. See Figure 3 for details of this timing. Contemporaneous penetration and leads of penetration serve as placebos that allow for tests for pre-trends within the county. The data include penetration from 2006 through 2011 and market risk from 2007 through 2011. The inclusion of leads and lags determines the available panel years, listed in the header for each column. Observations are county  $\times$  years. Controls are as described in Table 2. Standard errors in parentheses are clustered at the county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table 4:** 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 X Year Demographics			X
Mean of Dep. Var.	0.485	0.485	0.485
Observations	15,640	15,640	15,640

**Note:** The table shows coefficients from difference-in-differences regressions in which the dependent variable is the average *demographic* risk score in the market (county). Demographic risk scores are calculated by the authors using data from the Medicare Beneficiary Summary File, and are based on age, gender, and Medicaid status, but *not diagnoses*. Observations are county  $\times$  years. Controls are as described in Table 2. Standard errors in parentheses are clustered at the county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table 5: 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 X 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:** The table shows coefficients from difference-in-differences regressions in which the dependent variables are the mortality rate (columns 1 through 3) and the cancer incidence rate (columns 4 through 6), both calculated among county residents with age  $\geq 65$ . The mortality rate is derived using data from the National Center for Health Statistics. The cancer incidence rate data come from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute, which tracks cancer rates independently from rates observed in claims data. The smaller sample size in columns 4 through 6 reflects the incomplete geographical coverage of SEER cancer incidence data. Observations are county  $\times$  years. Controls are as described in Table 2. Standard errors in parentheses are clustered at the county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table 6:** The Role of Insurer-Provider Integration: Effects by Contract Type

		Dependent Variable: County-Level Average Risk Score		
		(1)	(2)	(3)
HMO Penetration	t (placebo)	0.011 (0.033)	0.035 (0.036)	0.026 (0.035)
PPO Penetration	t (placebo)	-0.006 (0.033)	0.011 (0.038)	0.012 (0.037)
PFFS Penetration	t (placebo)	-0.015 (0.032)	0.000 (0.037)	0.001 (0.036)
HMO Penetration	t-1	0.137** (0.027)	0.098** (0.028)	0.101** (0.028)
PPO Penetration	t-1	0.115** (0.026)	0.072* (0.028)	0.068* (0.028)
PFFS Penetration	t-1	0.048* (0.023)	0.063* (0.025)	0.058* (0.025)
Main Effects				
County FE		X	X	X
Year FE		X	X	X
Additional Controls				
State X Year Trend			X	X
County X Year Demographics				X
Observations		15,640	15,640	15,640

**Note:** The table shows coefficients from difference-in-differences regressions in which the dependent variable is the average risk score in the market (county). Effects of contemporaneous ( $t$ ) and lagged ( $t - 1$ ) Medicare Advantage (MA) penetration are displayed, disaggregated by the category of MA plan (HMO/PPO/PFFS) experiencing the net change in enrollment. Because MA risk scores are calculated using diagnosis data from the prior plan year, changes in MA enrollment can plausibly affect reported risk scores via differential coding only with a lag. Thus, contemporaneous penetration serves as a placebo that allows for a test for pre-trends within the county. Observations are county  $\times$  years. Controls are as described in Table 2. Regressions additionally control for penetration of all other contract types, so that the net change in penetration summed across contract types equals the “MA penetration” variable in Table 2. Standard errors in parentheses are clustered at the county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table 7:** The Role of Insurer-Provider Integration: Effects by Firm Ownership

		Dependent Variable: County-Level Average Risk Score		
		(1)	(2)	(3)
Non-Physician-Owned Plan Penetration	t (placebo)	0.114** (0.036)	0.043 (0.039)	0.041 (0.039)
Physician-Owned Plan Penetration	t (placebo)	0.007 (0.015)	-0.002 (0.019)	-0.002 (0.019)
Non-Physician-Owned Plan Penetration	t-1	0.056** (0.011)	0.064** (0.012)	0.061** (0.011)
Physician-Owned Plan Penetration	t-1	0.125** (0.031)	0.156** (0.033)	0.156** (0.031)
Main Effects				
County FE		X	X	X
Year FE		X	X	X
Additional Controls				
State X Year Trend			X	X
County X Year Demographics				X
Observations		15,640	15,640	15,640

**Note:** The table shows coefficients from difference-in-differences regressions in which the dependent variable is the average risk score in the market (county). Effects of contemporaneous ( $t$ ) and lagged ( $t - 1$ ) Medicare Advantage (MA) penetration are displayed, disaggregated by whether plans were physician-owned, following the definitions constructed by [Frakt, Pizer and Feldman \(2013\)](#) (see Section [A.4](#) for full details). Because MA risk scores are calculated using diagnosis data from the prior plan year, changes in MA enrollment can plausibly affect reported risk scores via differential coding only with a lag. Thus, contemporaneous penetration serves as a placebo that allows for a test for pre-trends within the county. Observations are county  $\times$  years. Controls are as described in Table 2. Standard errors in parentheses are clustered at the county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .



**Table 8:** Alternative Identification: Coding Differences at Age 65 Threshold, MA vs. FFS

	Dependent Variable:					
	Risk Score		Count of HCCs		At least 1 HCC	
	(1)	(2)	(3)	(4)	(5)	(6)
Post-65	0.023* (0.011)	0.023** (0.007)	0.031 (0.023)	0.031* (0.015)	0.005 (0.010)	0.005 (0.007)
Selected MA	-0.072** (0.013)		-0.175** (0.027)		-0.071** (0.014)	
Post-65 X Selected MA	0.044* (0.021)	0.044** (0.015)	0.123** (0.043)	0.123** (0.030)	0.060** (0.020)	0.060** (0.015)
Person FE		X		X		X
Mean of Dep. Var.	0.55	0.55	0.57	0.57	0.33	0.33
Observations	12,142	12,142	12,142	12,142	12,142	12,142

**Note:** The table shows coefficients from difference-in-differences regressions in which the dependent variables are the average risk score (columns 1 and 2), the count of chronic conditions (HCCs) used in the risk adjustment formula (columns 3 and 4), and an indicator for being diagnosed with any condition used in the risk adjustment formula (columns 5 and 6). The regressions compare these coding outcomes across the age 65 threshold among consumers who select MA vs. FFS, pre- and post-entering Medicare. Data are from the Massachusetts All-Payer Claims Dataset. Pre-65 claims are from commercial/employer plans. Post-65 claims are from Medicare Advantage plans for MA enrollees and Medigap plans for FFS enrollees. The sample is restricted to individuals who join FFS or MA within one month of their 65th birthday and who have at least 6 months of continuous coverage before and after their 65th birthday. Odd columns control for gender, region of Massachusetts, and month of enrollment. Even columns include individual fixed effects. Data are described more thoroughly in Section A.6. Observations are person-years. Standard errors in parentheses are clustered at the person level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table 9:** Counterfactual: Implied MA Enrollment Effects of Removing Coding Subsidy

Study	Estimated semi-price elasticity of demand	Implied semi-payment elasticity of demand	Implied enrollment effect of removing overpayment due to coding	
			Relative to counterfactual of no CMS coding adjustment (6.4% reduction in payments)	Relative to counterfactual of 3.4% coding deflation by CMS (3% reduction in payments)
Cabral, Geruso, and Mahoney (2014)	-0.0068	-0.0034	-17%	-8%
Atherly, Dowd, and Feldman (2003)	-0.0070	-0.0035	-18%	-8%
Town and Liu (2003)	-0.0090	-0.0045	-23%	-11%
Dunn (2010)	-0.0129	-0.0065	-33%	-15%

**Note:** Table displays back-of-the-envelope calculations for MA market segment enrollment under a counterfactual in which the subsidy to differential coding is removed. A monthly benchmark (i.e., plan payment) of \$800 is assumed. Price semi-elasticities are taken from the prior literature. Implied semi-payment elasticities of demand have been derived from semi-price elasticities assuming a pass-through rate of 50%. See text for full details.

## 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 levels of these county-specific payments are tied to historical fee-for-service (FFS) Medicare costs in the county. County capitation rates were originally intended to capture the cost of enrolling the “national average beneficiary” in the FFS 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 in MA, relying primarily on demographics.<sup>67</sup> 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  $R_{ijc} = \phi_{jc} \cdot x_{ic} \Lambda$ , 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  $\phi_{jc}$  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. These bids 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 the benchmark, which the plan is required to fold 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. This implies that any estimation of coding subsidies should be based on the capitation payment to plans inclusive of any “rebate.”

### A.2 Linearity of Market-level Average Risk Curve

Throughout much of Section 3 we impose the assumption that coding differences across plans can be represented by the difference between two additive plan-dependent coding factors. Under this assumption, the slope of the average risk curve in the market is exactly equal to the average coding difference in the population. This relationship ( $\frac{\partial \bar{r}}{\partial \theta} = \alpha_k - \alpha_j$ ) also holds under the weaker assumption that any individual-specific heterogeneity in the plan-dependent component of the risk score is orthogonal to plan  $k$ ’s market share ( $\theta^k$ ). In that case the slope of the average risk curve in the market is exactly equal to the average coding difference across plans.

To formalize this statement, let individual  $i$ ’s risk score in plan  $j$  be equal to  $r_{ij} = \hat{r}_i + \alpha(\gamma_j, \psi_j)$ , and let  $i$ ’s risk score in plan  $k$  be equal to  $r_{ik} = \hat{r}_i + \alpha(\gamma_k, \psi_k) + \epsilon_{ik}$  where  $\epsilon_{ik}$  has mean  $\bar{\epsilon} = 0$  and represents individual-specific heterogeneity in the relative coding intensity between plans  $j$  and  $k$ . For individual  $i$ , this generates a differential risk score of  $\alpha(\gamma_k, \psi_k) + \epsilon_{ik} - \alpha(\gamma_j, \psi_j)$ . This produces an average upcoding factor of  $\frac{1}{N} \sum (\alpha(\gamma_k, \psi_k) + \epsilon_{ik} - \alpha(\gamma_j, \psi_j)) = \alpha(\gamma_k, \psi_k) - \alpha(\gamma_j, \psi_j)$ .

Letting  $\mathbb{1}[k_i(\theta)]$  represent the indicator function for choosing  $k$ , market average risk can be expressed as  $\bar{r} = \frac{1}{N} \sum (\hat{r}_i + \alpha(\gamma_j, \psi_j) + \mathbb{1}[k_i(\theta)](\alpha(\gamma_k, \psi_k) + \epsilon_{ik} - \alpha(\gamma_j, \psi_j)))$ . Assuming that  $\epsilon_k$  is orthogonal to  $\theta$ , differentiating  $\bar{r}$  with respect to  $\theta$  produces:

$$\frac{\partial \bar{r}}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{1}{N} \sum (\hat{r}_i + \alpha(\gamma_j, \psi_j) + \mathbb{1}[k_i(\theta)](\alpha(\gamma_k, \psi_k) + \epsilon_{ik} - \alpha(\gamma_j, \psi_j))) = \alpha(\gamma_k, \psi_k) - \alpha(\gamma_j, \psi_j). \quad (12)$$

<sup>67</sup>From 2001-2003 inpatient diagnoses were used in risk adjustment, but in order to weaken insurer incentives to send more enrollees to the hospital, CMS only gave these diagnoses a 10% weight in the risk adjustment payment, leaving the other 90% of the payment based on demographics only.

Therefore, in the case of individual-level heterogeneity in “upcodability,” the slope of the market-level risk curve still identifies the mean coding difference across plans or market segments.

### A.3 Estimates of Selection

Section 5 describes the results of the main analysis in which we regress county-level averages of risk scores on lagged MA penetration in the county to estimate coding differences. For completeness, here we estimate selection on risk scores, using an analogous set of regressions. Under the assumption that MA penetration changes are exogenous to changes in underlying population health conditional on our controls, selection on risk scores can be estimated by regressing either the average risk score within FFS or the average risk score within MA on contemporaneous and lagged penetration. Note that this reveals only compensated selection, not uncompensated selection as estimated in [Brown et al. \(2014\)](#) and [Cabral, Geruso and Mahoney \(2014\)](#).

Table A1 presents the selection results. Coefficients on contemporaneous MA penetration identify pure selection effects. If selection were monotonic (such as in Figure 1), then positive contemporaneous coefficients in both markets would indicate that as penetration increased, the marginal beneficiary choosing MA was high risk relative to the MA average and low risk relative to the FFS average (according to prior-year diagnoses), increasing the average risk score in both pools. In Table A1, estimates for both FFS and MA risk are imprecise, yielding confidence intervals consistent with a broad range of selection effects, including the findings in [Newhouse et al. \(2012\)](#) of advantageous selection into MA of 4 to 9% of the risk score in 2008.

An important component of selection effects may be captured by the lagged penetration coefficient: Research on MA enrollment by [Sinaiko, Afendulis and Frank \(2013\)](#) shows that the majority of new MA enrollees are age 65, implying that most of the shift in MA penetration is likely occurring among the newly Medicare-eligible. In Table A1, this would cause a significant fraction of selection effects to be captured by the lagged coefficient, as new MA enrollees aren’t assigned diagnosis-based risk scores until their second year. However, interpreting selection effects in Table A1 is difficult because coefficients on lagged MA penetration are affected by: (i) selection on risk score trajectory and (ii) selection on the unobserved contemporaneous risk score for new enrollees who are not assigned a diagnosis-based score until their second year.

It is important to note that unlike these within-market-segment results, the regressions comprising our main analysis, which examine the effect of lagged penetration on *overall* county risk, are unaffected by selection and yield a straightforward identification of pure coding effects.

### A.4 Supplemental Analysis on Plan Ownership

In Section 5 we described results that identified heterogeneity in coding practices across plans with different levels of insurer-provider integration. For our results in Table 7, we calculated MA penetration separately for physician-owned plans, using data constructed by [Frakt, Pizer and Feldman \(2013\)](#). Here, we describe those data and results in more detail.

[Frakt, Pizer and Feldman \(2013\)](#) gathered data on provider ownership of plans via plan websites and governance documents for plan year 2009. They limited attention to coordinated care MA-Part D plans (e.g. HMOs and PPOs), excluding employer plans, PFFS plans, and MA plans without drug coverage. We apply their integration flag to our data covering years 2006-2011, using publicly available CMS plan crosswalk files to link the 2009 plan IDs across years. The restriction of [Frakt, Pizer and Feldman \(2013\)](#) to exclude non-drug plans from classification and our implicit assumption that physician ownership was constant from 2006 to 2011 could introduce measurement error, which would bias against our finding of a difference in coding between plans classified as physician-owned and not.

## A.5 The Role of Electronic Health Record (EHR) Adoption

Regressions in Table A5 analyze the extent of coding intensity differences across markets classified by differences in the local adoption of electronic health records (EHRs). Here we describe the data used to classify local markets by EHR adoption.

CMS, in cooperation with the Office of the National Coordinator for Health Information Technology, has collected data on meaningful use of EHR systems within physician office settings at the county level. Since 2011, physician offices serving Medicare patients have been incentivized with financial bonuses to report on meaningful EHR use to CMS. We use reports of EHR use during the first year of the incentive program (2011) as a proxy for the existing local EHR infrastructure during our sample period (2006-2011). Within each county, we normalize the count of physicians reporting office EHR adoption by the county Medicare population. Then we define an indicator for high EHR adoption by splitting this metric at the median. Interaction terms in Table A5 between lagged penetration and this indicator for high EHR adoption yield coefficients very close to zero.

## A.6 Data Used in Section 6

For the analysis in Section 6, we obtained data on Medicare Advantage and Medigap claims from the new Massachusetts All-Payer Claims Dataset (Mass APCD). The Mass APCD is maintained by the Massachusetts Center for Health Information and Analysis (CHIA). The database includes the universe of health insurance claims for individuals receiving medical services under a commercial plan 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 generated by the state using Social Security numbers, names, and addresses. This ID is available only for 2011-12.

We identify two groups of individuals in the Mass APCD. The first group consists of all individuals enrolling in a product identified as a Medicare Advantage plan within one month of their 65th birthday. We identify Medicare Advantage plans using an identifier provided by CHIA. We verify that the CHIA identifier for Medicare Advantage products is accurate by matching the names of the payers in the Mass APCD data to publicly available Medicare Advantage enrollment data provided by CMS. We also check the age distribution of enrollees in these plans to ensure that there is a discontinuous spike in the density at age 65.

The second group consists of all individuals enrolling in a product identified as a Medigap plan within one month of their 65th birthday. We identify Medigap plans using an identifier provided by CHIA. Again, we verify that the CHIA identifier for Medigap plans is accurate by matching the names of the payers in the Mass APCD data to publicly available information about the insurers competing in the Massachusetts Medigap market. Additionally, we observe the portion of total spending paid by Medicare. For almost all of the plans identified as Medigap plans, this value was between 70% and 90%, confirming that these products are Medigap plans.

For both groups, we identify the subset of individuals with some form of coverage for every month of 2011 and 2012. We drop any individuals with some form of *Medicare* coverage prior to joining MA or Medigap at 65. This removes, for example, Medicare enrollees who gain coverage via disability rather than automatically becoming eligible at age 65. Because Medicaid data is excluded from our version of the Mass APCD, this results in a sample of individuals with continuous commercial/employer coverage prior to their 65th birthday and continuous Medigap/MA coverage after their 65th birthday. To ensure that we have enough claims data to calculate risk scores, we restrict the sample to individuals with at least six months of data prior to joining Medigap/MA and six months of data after joining Medigap/MA. This effectively eliminates anyone joining Medigap/MA before July 2011 and after July 2012.

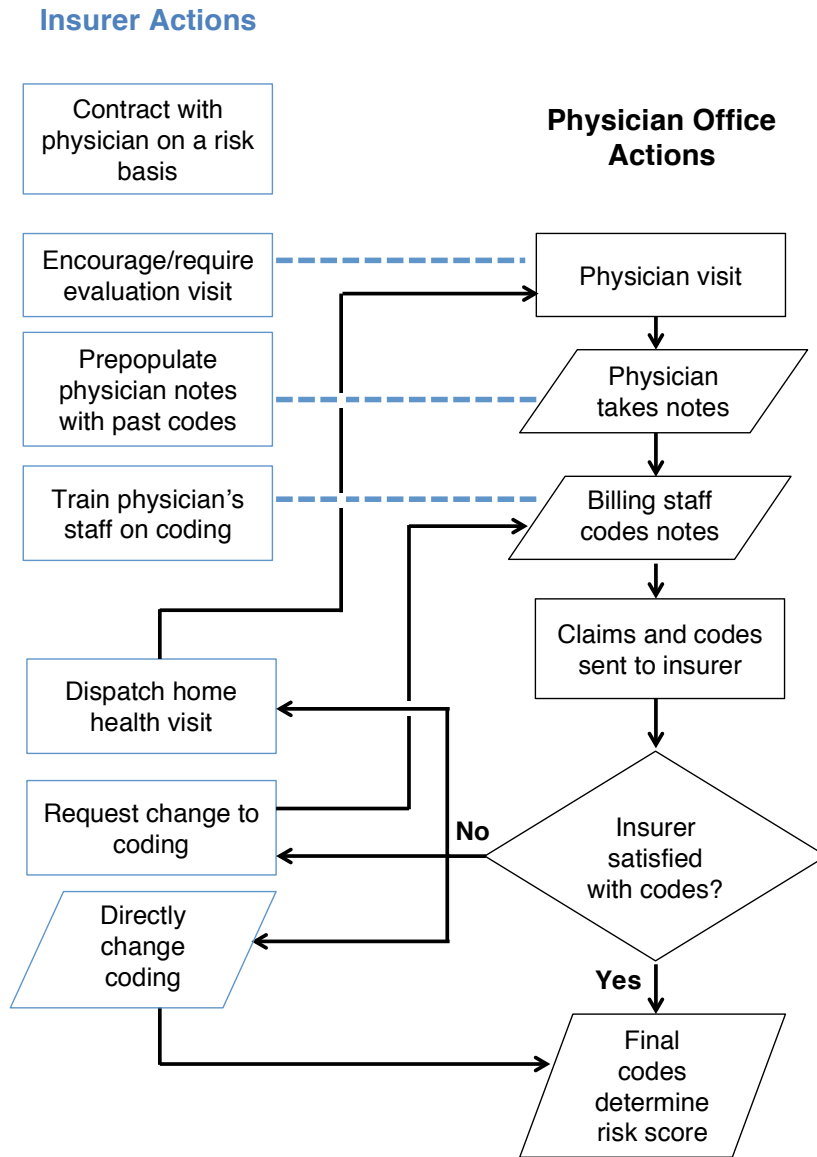
For the resulting set of individuals, we restrict attention to claims from medical plans only, excluding prescription drug plans.<sup>68</sup> We then assign individuals to cohorts based on the month in which they joined Medigap/MA. We specify separate “pre-65” and “post-65” periods for each cohort. We require that the pre-65 and post-65 periods be of equal length and that they consist of the same months from different years. Given these restrictions, we choose the longest possible period for each cohort. For example, for the individuals joining Medigap/MA in July 2011, we specify the pre-65 period to be January-June 2011 and the post-65 period to be January-June 2012. For the individuals joining Medigap/MA in February 2012, we specify the pre-65 period to be February-December 2011 and the post-65 period to be February-December 2012.<sup>69</sup>

For each individual, we calculate a pre-65 and a post-65 risk score. The pre-65 (post-65) risk score is calculated by identifying all diagnoses from medical claims incurred during the pre-65 (post-65) period and running those diagnoses through the 2011 version of the CMS-HCC SAS risk scoring program. The program is freely available from CMS. It maps each individual’s diagnoses to a set of 70 chronic conditions. It then multiplies indicators for these chronic conditions by a set of weights estimated by CMS using claims data from fee-for-service Medicare. The product of the individual’s chronic conditions indicators and the weights is the risk score. We also calculate each individual’s number of chronic conditions and construct a dummy variable indicating whether the individual has at least one chronic condition.

---

<sup>68</sup>The CMS-HCC risk adjustment model used for MA payments is based only on diagnoses from medical claims.

<sup>69</sup>Alternatively, we could have required that the pre-65 and post-65 periods be contiguous. For example, for individuals joining Medigap/MA in July 2011, we could specify the pre-65 period to be January-June 2011 and the post-65 period to be July-December 2011. This alternative specification produces results nearly identical to those produced by the specification outlined above.

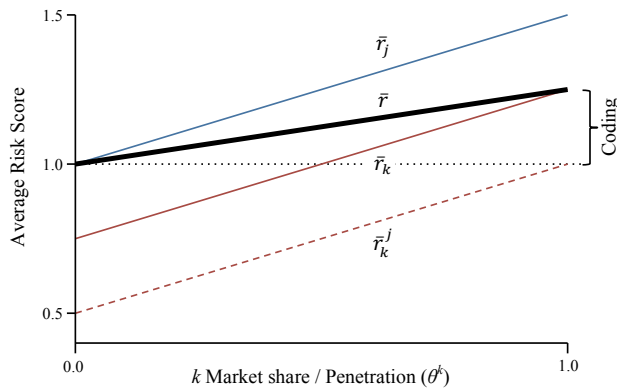
**Figure A1: How Risk Scores are Influenced by Insurers**

**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 the actions of the provider's billing and coding staff, are towards the right in black boxes. Actions that can immediately result in generation of a code are represented by rhombuses.

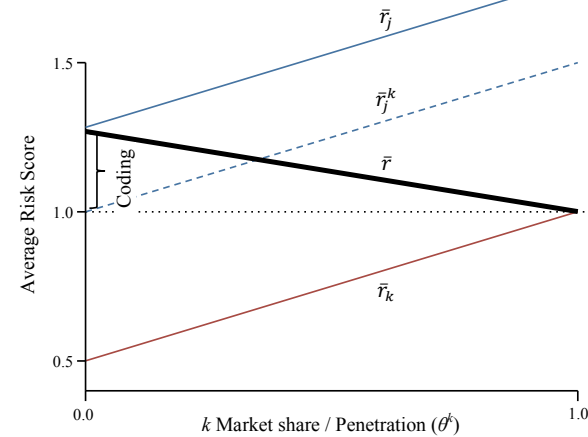


**Figure A2: Identifying Coding Differences in Selection Markets: Alternative Forms of Selection**

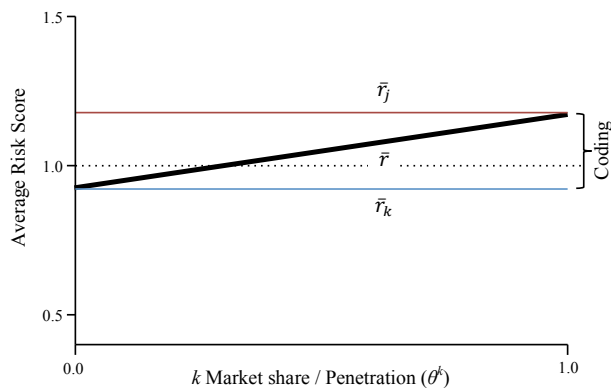
(A) Advantageously Selected Plan Codes More Intensely



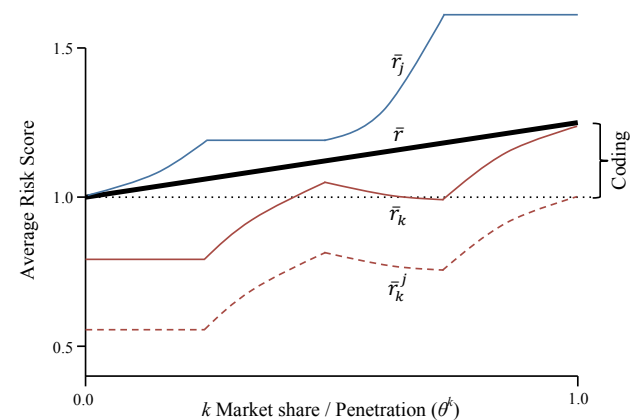
(B) Advantageously Selected Plan Codes Less Intensely



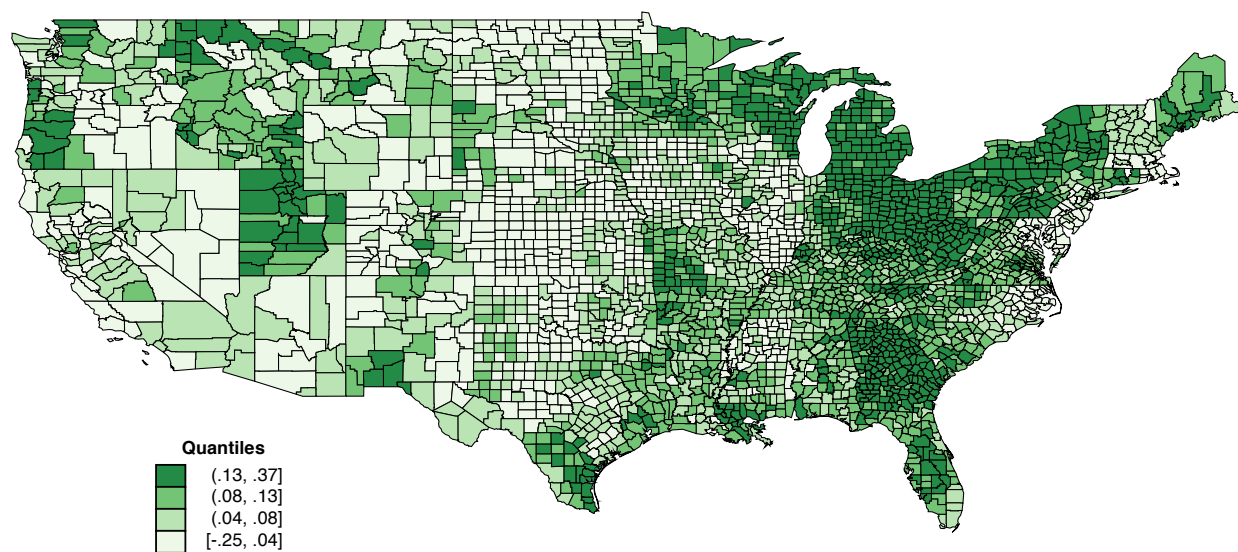
(C) No Selection and Differential Coding Intensity



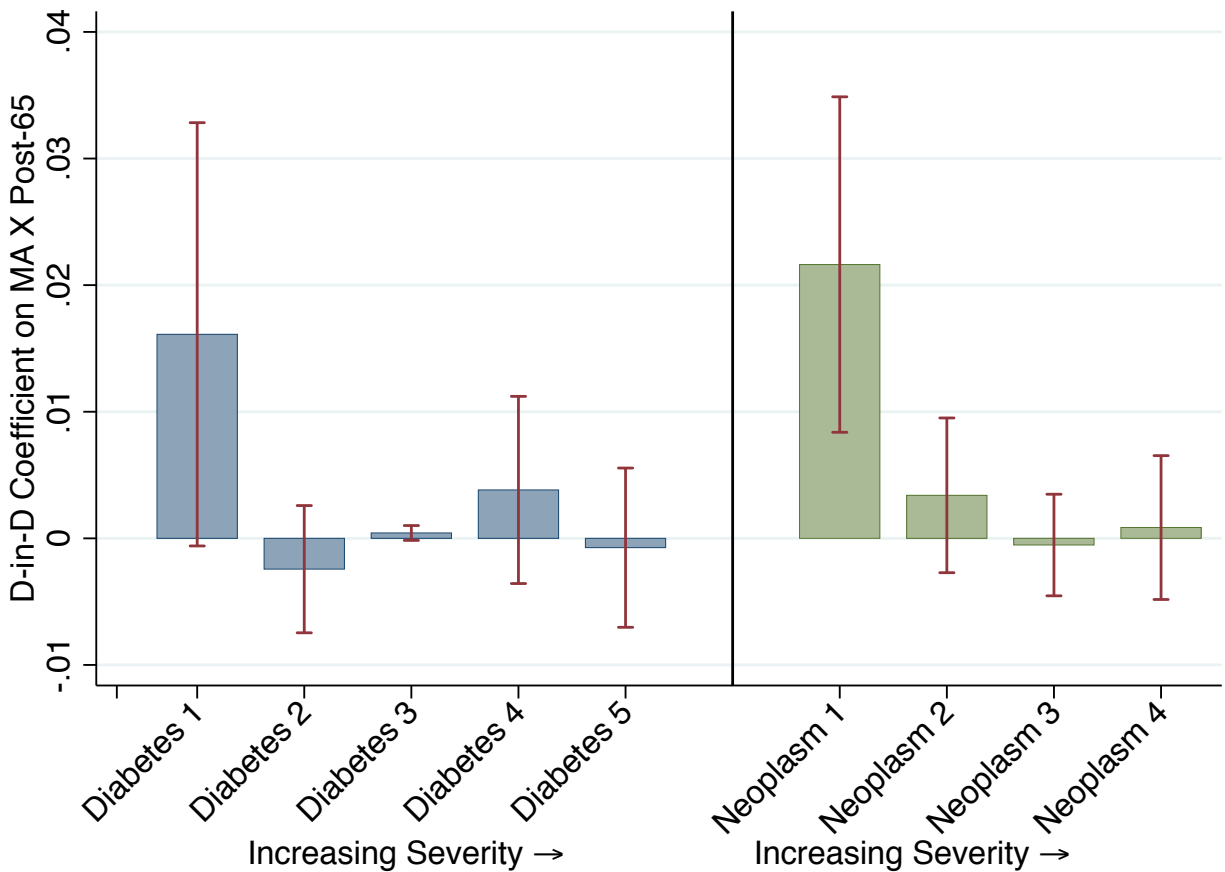
(D) Selection is Nonlinear and Non-monotonic



**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  $k$ ,  $\theta^k$ . The vertical axis measures the average risk score: Average risk in  $j$  is  $\bar{r}_j$ , average risk in  $k$  is  $\bar{r}_k$ , 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  $k$  enrollees would have been assigned under segment  $j$  coding practices,  $\bar{r}_k^j$ . All consumers choose either plan  $j$  or plan  $k$ . If and only if there are coding differences between  $j$  and  $k$ , then the slope of the market-level risk curve with respect to marketshare ( $\frac{\partial \bar{r}}{\partial \theta^k}$ ) will be different from zero.

**Figure A3: Geography of Growth in Medicare Advantage, 2006 to 2011**

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

**Figure A4:** Coding Differences by Severity Across the Age 65 Threshold, MA vs. FFS

**Note:** The figure shows coefficients from difference-in-differences regressions in which the dependent variable is an indicator for being diagnosed with a specific level of diabetes or cancer severity. These regressions compare pre and post diagnoses among consumers aging into Medicare and selecting MA versus FFS. Each bar represents the coefficient on  $\text{Post-65} \times \text{Selected MA}$  from a separate OLS regression. The specification parallels the regressions in Table 8. Red bars indicate the 95% confidence intervals. Data are from the Massachusetts All-Payer Claims Database, as described in the text. Prior to age 65, all consumers in the sample are continuously enrolled in commercial/employer plans. We describe the construction of these statistics in more detail in Appendix A.6.

**Table A1:** Selection Results: Effects on within-FFS and within-MA Risk Scores

	Dependent Variable:					
	Mean FFS Risk Score			Mean MA Risk Score		
	(1)	(2)	(3)	(4)	(5)	(6)
MA Penetration t	0.037 (0.026)	0.040 (0.034)	0.040 (0.033)	0.025 (0.062)	-0.024 (0.085)	-0.013 (0.083)
MA Penetration t-1	0.045** (0.013)	0.030* (0.012)	0.026* (0.012)	0.087* (0.040)	0.116** (0.040)	0.130** (0.041)
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 X Year Demographics			X			X
Dep var mean	1.006	1.006	1.006	0.959	0.959	0.959
Observations	15,640	15,640	15,640	15,616	15,616	15,616

**Note:** The table shows coefficients from difference-in-differences regressions in which the dependent variables are the average FFS risk score in the county (columns 1 through 3) and the average MA risk score in the county (columns 4 through 6). Both contemporaneous and lagged coefficients represent tests of selection. Observations are county  $\times$  years. Controls are as described in Table 2. Standard errors in parentheses are clustered at the county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table A2:** Robustness to Alternative Subsets of the Identifying Variation

	Dependent Variable: County-Level Average Risk Score								
	Main Results			Identified by MA-Part D Penetration Changes			Identified by MA-non-Part D Penetration Changes		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MA Penetration $t$ (placebo)	0.007 (0.015)	0.001 (0.019)	0.001 (0.019)	0.011 (0.015)	0.008 (0.018)	0.007 (0.018)	-0.008 (0.024)	-0.012 (0.033)	-0.011 (0.033)
MA Penetration $t-1$	0.069** (0.011)	0.067** (0.012)	0.064** (0.011)	0.081** (0.014)	0.065** (0.016)	0.064** (0.016)	0.040** (0.013)	0.055** (0.019)	0.050** (0.018)
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 X Year Demographics			X			X			X
MA-non-Part D Penetration				X	X	X			
MA-Part D Penetration							X	X	X
Mean of Dep. Var.	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Observations	15,640	15,640	15,640	15,640	15,640	15,640	15,640	15,640	15,640

**Note:** The table shows coefficients from difference-in-differences regressions in which the dependent variable is the average risk score in the market (county). Effects of contemporaneous ( $t$ ) and lagged ( $t - 1$ ) Medicare Advantage (MA) penetration are displayed. Columns 1 through 3 reproduce Table 2 for ease of comparison. Columns 4 through 6 control for changes in MA penetration among MA plans not offering Part D drug benefits. Columns 7 through 9 control for changes in MA penetration among MA plans offering Part D drug benefits. Additional controls are as described in Table 2. Observations are county  $\times$  years. Standard errors in parentheses are clustered at the county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table A3:** Extended Placebo Tests: Effects of Contemporaneous Penetration and Leads

	Dependent Variable: County-Level Average Risk Score							
	available panel years:							
	2007-2011 (1)	2007-2010 (2)	2007-2009 (3)	2008-2011 (4)	2008-2010 (5)	2008-2009 (6)	2009-2011 (7)	2009-2010 (8)
MA Penetration $t+2$ (placebo)			0.044+ (0.023)			0.030 (0.036)		
MA Penetration $t+1$ (placebo)		0.017 (0.025)	0.032 (0.056)		-0.005 (0.015)	-0.019 (0.042)		-0.004 (0.034)
MA Penetration $t$ (placebo)	0.001 (0.019)	-0.021 (0.028)	-0.064 (0.071)	0.006 (0.017)	0.003 (0.025)	-0.025 (0.091)	0.011 (0.016)	0.014 (0.043)
MA Penetration $t-1$	0.064** (0.011)	0.076** (0.018)	0.084** (0.022)	0.041** (0.015)	0.038+ (0.022)	0.025 (0.038)	0.037 (0.032)	0.052 (0.090)
MA Penetration $t-2$				0.046* (0.022)	0.054* (0.024)	0.048 (0.041)	0.052+ (0.031)	0.100 (0.061)
MA Penetration $t-3$							0.023 (0.024)	-0.033 (0.039)
Main Effects								
County FE	X	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X	X
Additional Controls								
State X Year Trend	X	X	X	X	X	X	X	X
County X Year Demographics	X	X	X	X	X	X	X	X
Observations	15,640	12,512	9,384	12,512	9,384	6,256	9,384	6,256

**Note:** The table shows coefficients from difference-in-differences regressions in which the dependent variable is the average risk score in the market (county). Effects of future ( $t + 2$ ,  $t + 1$ ), contemporaneous ( $t$ ), and lagged ( $t - 1$ ,  $t - 2$ ,  $t - 3$ ) Medicare Advantage (MA) penetration are displayed. Because MA risk scores are calculated using diagnosis data from the prior plan year, changes in MA enrollment can plausibly affect reported risk scores via differential coding only with a lag. See Figure 3 for details of this timing. Contemporaneous penetration and leads of penetration serve as placebos that allow for tests for pre-trends within the county. The data include penetration from 2006 through 2011 and market risk from 2007 through 2011. The inclusion of leads and lags determines the available panel years, listed in the header for each column. Observations are county  $\times$  years. Controls are as described in Table 2. Standard errors in parentheses are clustered at the county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table A4:** Falsification Test: Effects on Medicare Age Distribution

	Dependent Variable:		Dependent Variable:			
	Fraction ≥65	Indicator for Age Bin, Conditional on ≥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 X Year Demographics						
Observations	15,640	15,640	15,640	15,640	15,640	15,640

**Note:** The table shows coefficients from difference-in-differences regressions in which the dependent variables are indicators for age ranges. The dependent variable in column 1 is the fraction of the Medicare population with age  $\geq 65$ . The dependent variables in columns 2 through 6 are the fractions of the Medicare population in the indicated age bins, conditional on age  $\geq 65$ . Data on the Medicare age distribution come from the Medicare Beneficiary Summary File. Observations are county  $\times$  years. Controls are as described in Table 2. Standard errors in parentheses are clustered at the county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table A5:** No Coding Interaction with Electronic Health Records

	Dependent Variable: County-Level Average Risk Score		
	(1)	(2)	(3)
MA Penetration $t$	-0.016 (0.026)	-0.024 (0.029)	-0.020 (0.029)
MA Penetration $t-1$	0.069** (0.016)	0.069** (0.017)	0.066** (0.016)
High EHR X MA Penetration $t$	0.042 (0.028)	0.051 (0.028)	0.043 (0.027)
High EHR X MA Penetration $t-1$	-0.002 (0.018)	-0.005 (0.017)	-0.006 (0.017)
Main Effects			
County FE	X	X	X
Year FE	X	X	X
Additional Controls			
State X Year Trend		X	X
County X Year Demographics			X
Observations	15,640	15,640	15,640

**Note:** The table shows coefficients from difference-in-differences regressions in which the dependent variable is the average risk score in the market (county). Effects of contemporaneous ( $t$ ) and lagged ( $t - 1$ ) Medicare Advantage (MA) penetration are displayed. All regressions include interactions between the MA penetration variables and an indicator for high electronic health record (EHR) adoption by physician offices in the county. Data on EHR adoption were assembled by CMS and the Office of the National Coordinator for Health Information Technology (see Section A.5 for full details). Observations are county  $\times$  years. Controls are as described in Table 2. Standard errors in parentheses are clustered at the county level. +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ .