Upcoding or Selection?

Evidence from Medicare on Squishy Risk Adjustment*

Michael Geruso† Timothy Layton‡

December 17, 2014

Abstract

Risk adjustment is commonly used in health insurance markets to deal with problems of adverse selection and cream skimming by compensating health plans for insuring consumers whose diagnoses imply high expected costs. However, in all real world risk adjustment systems, insurers themselves report the diagnoses that determine enrollee risk scores and ultimately insurer payments. If risk scores are manipulable, insurers that "upcode" enrollees will extract higher payments. We model upcoding in the presence of adverse selection and develop a novel strategy for separately identifying upcoding from selection in data. We apply this strategy to analyze diagnosis coding by Medicare Advantage insurers. We find that enrollees in Medicare Advantage plans generate 7% higher risk scores than what the same enrollees would generate under Traditional Medicare, where coding incentives are weaker. Absent a coding inflation correction, this implies excess payments to Medicare Advantage of around $11 billion annually and a distortion in seniors' choice between Medicare Advantage and Traditional Medicare. This choice distortion *worsens* with increasing insurer competition. We also find evidence that coding intensity is higher in plans with higher levels of insurer-provider integration, further distorting seniors' choices to more integrated plans.

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

†University of Texas at Austin and NBER. Email: mike.geruso@gmail.com

‡Harvard Medical School. Email: timothyjlayton@gmail.com

1 Introduction

Risk adjustment is widely used in health insurance markets to counteract inefficiencies caused by adverse selection. By modifying payments to insurers to compensate for enrolling a high expected cost consumer, risk adjustment weakens incentives for insurers to engage in cream-skimming—that is, inefficiently distorting insurance product characteristics to attract lower-cost enrollees—as in Rothschild and Stiglitz (1976).[1] In a setting where product characteristics are fixed, risk adjustment also weakens the link between a plan's price and the average cost of its enrollees, which could otherwise cause inefficient sorting or market unravelling (Akerlof, 1971).[2] While the intuition underlying risk adjustment is straightforward, the mechanism relies on an implicit assumption that the regulator can construct individual "risk scores" that summarize the consumer's expected cost to the insurer but do not depend on insurer actions. Because in all real world payment systems the insurers themselves report the diagnoses that determine enrollee risk scores, and therefore ultimately influence the insurer's payments from the regulator, it is plausible that this assumption does not hold in practice. Whether and to what extent risk scores can be influenced by insurers is of significant practical importance, as risk adjustment is the primary regulatory mechanism used to counteract distortions caused by asymmetric information in US health insurance markets, including in Medicare, some state Medicaid programs, and the Affordable Care Act Health Insurance Exchanges. Currently over 50 million Americans are enrolled in a risk-adjusted insurance market.

In this paper, we show that even when successful in counteracting the Rothschild and Stiglitz (1976) and Akerlof (1971) selection inefficiencies, risk adjustment introduces a new distortion by incentivizing intensive coding. We begin with a simple theoretical framework that illuminates how coding differences across insurers or across market segments impact total spending in public insurance programs like Medicare and Medicaid. We also characterize the consumer choice distortions that can arise in any risk-adjusted market, including the new Health Insurance Exchanges, as a result of the implicitly larger regulatory transfer or subsidy to plans with higher coding intensity. We then

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

[2] Empirical studies of the Rothschild and Stiglitz (1976) inefficiency in health insurance include Frank, Glazer and McGuire (2000); Buchmueller and DiNardo (2002); Cao and McGuire (2003); Carey (2014). Empirical studies of the Akerlof (1971) inefficiency include Einav, Finkelstein and Cullen (2010); Carlin and Town (2010); Lustig (2011); Bundorf, Levin and Mahoney (2012); Geruso (2012); Hackmann, Kolstad and Kowalski (2013); Handel (2013); Handel, Hendel and Whinston (2013); Layton (2014).

provide the first econometrically identified estimates of coding intensity differences across insurers. Specifically, we study coding differences between the uncoordinated reporting of diagnoses in the Traditional Medicare program (TM) and Medicare Advantage (MA), the private Medicare option in which premiums are heavily or fully subsidized by risk-adjusted capitation payments to private insurers.

Despite wide policy and research interest in "upcoding," the extent of coding differences across insurers or across the public and privatized arms of programs like Medicare and Medicaid is largely unknown. Estimates of upcoding are potentially confounded by adverse selection: An insurer might report an enrollee population with higher-than-average risk scores either because the consumers choosing its plan are in worse health (selection) or because for the same individual, the insurer's coding practices result in higher risk scores (upcoding). Because of this central identification problem, there has been little empirical work on coding intensity in any US health insurance market. Two important exceptions are Silverman and Skinner (2004) and Dafny (2005), which exploit changes during the 1990s in how TM compensated hospitals on the basis of certain diagnoses, showing that hospital coding patterns responded to track the pattern of reimbursement changes for particular diagnoses. The limitation of such difference-in-differences approaches around a change to a subset of codes is that while these studies clearly demonstrate that coding behavior is endogenous to the payment system (and for Dafny, 2005 without any change in real service provision), they cannot quantify the full impact of endogenous coding across all diagnostic areas, and therefore the overall impact on the market.

To identify overall coding intensity differences across insurers within a market, we develop a general method for separating upcoding from selection, which is applicable to any market where consumers sort endogenously. The core insight of our identification approach is novel, but straightforward: If the same individual would generate a different risk score under two insurers and if we observe a change in market share of the two insurers, then we should also observe changes to the observed *market-level* average of the risk score. Such a pattern could not be rationalized by selection, because selection can affect only the sorting of risk types across insurers within the market, not the overall market-level distribution of reported risk.[3] Our strategy is related to that of Chetty,

[3]With this method, the only assumption required to ensure an unbiased estimate of the existence and extent of upcoding is exogeneity of the changes in market share to true underlying population health. Our approach contrasts with past attempts by regulators and researchers to identify upcoding by following individual switchers across insurers and observing how risk scores change after the switch. Such a method requires assuming that consumers change health plans

Friedmand and Rockoff (2014), who identify teacher value-added via changes to the composition of teaching staff within a grade over time, avoiding biases introduced by student sorting by examining changes in outcomes averaged over the entire grade level, rather than at the teacher level.[4]

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

We show that it is difficult to rationalize our results by the alternative explanation that true population health was changing contemporaneously with these penetration changes within markets. First, the effect is large: a 7% increase in the average risk score is equivalent to 6% of all consumers in the market becoming paraplegic, 15% of all consumers contracting HIV, or 58% becoming diabetics.[5] Second, we 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 that are consistent with our findings: We find that risk scores respond to MA penetration changes with a one-year lag, showing no response to contemporaneous changes. Risk scores also show no response to penetration changes in future years, supporting the parallel trends assumption of our strategy. Further, because a portion of the risk scoring algorithm is based on demographic characteristics that are not manipulable by insurers, we verify in a separate falsification test that changes in MA penetration predict only changes to the diagnosis-based portion of the risk score and not to the demographic portion. Finally, we show that at the county level MA penetration does not predict other observable time-varying county

for reasons unrelated to their health. See for example, Government Accountability Office (2013). Our method does not require this identifying assumption. Additionally, results from studies of switchers, unlike results from our method, are often not generalizable to the entire population, prohibiting estimation of the overall effect of coding intensity differences on the public's finances.

[4] A closer analog of our 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.

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

characteristics that indicate health status but were not plausibly affected by insurer coding practices, including external measures of cancer morbidity, overall mortality, and the age distribution inside and outside of Medicare.

In addition to estimating the overall coding differences between TM and MA, we identify coding heterogeneity within MA across plan types. Because diagnosis codes ultimately originate from provider visits, insurers face a principal-agent problem in contracting with physicians. Insurers have full knowledge of the risk-scoring algorithms that convert diagnosis codes to insurer payments, but their ability to influence *providers* to adhere to coding practices that maximize *insurer* profits may vary by contract type. For example, while HMOs can create risk-based contracts with physician groups that fully pass-through coding incentives, Private Fee-for-Service plans do not have access to this (potentially effective) tool for influencing provider coding practices. We build on our main finding by investigating the degree to which the level of insurer-provider integration influences a plan's ability to induce intensive coding. We show that coding intensity is significantly higher for managed care MA plans, which are likely to have closer ties to physician groups, than for insurers with less provider integration. Risk scores in MA HMO plans are about 10% larger than they would have been for the same beneficiary in TM, while risk scores in Private Fee-for-Service plans are only 4-5% larger.[6] The practical importance of this heterogeneity is that an across-the-board deflation of MA risk scores, similar to the policy introduced by Medicare in 2010, is unlikely to be the optimal policy to counter the distortions introduced by differential coding.

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

Our model shows how differences in coding may cause excess public spending and always cause transfers across health plans that distort consumers' choices—in our case between the market seg-

[6] Employer-sponsored MA plans, which have weaker incentives for upcoding due to an additional principal-agent problem, also display weaker coding behavior.

ments of Traditional Medicare and Medicare Advantage. A non-obvious result that emerges from our modeling is that for many policy questions regarding the public finance of health insurance and regulatory incidence, it is not necessary to take a stand as to which insurer's coding regime is objectively correct. In our empirical setting, this means that it doesn't matter whether physicians under TM pay too little attention to coding or whether MA insurers pay too much attention to coding. This implies that the fine line between squishy legal coding and illegal "upcoding," which has been a focal point for regulators, lacks economic significance. Only relative differences in coding intensity matter for consumer choice and public spending.[7]

Second, we provide the first econometric evidence of upcoding in any private insurance market, which has implications both for public costs and consumer choice distortions. Medicare is the costliest public health insurance program in the world and makes up a significant fraction of US government spending. In 2014, the Medicare Advantage program accounted for 30% of total Medicare spending. Absent a coding correction, our estimates imply excess payments of around $11.4 billion to Medicare Advantage plans annually, or about $700 per MA enrollee per year. To put the magnitude in context, this is about twice as large as the Brown et al. (2014) estimate of the increase in excess payments to MA plans due to uncompensated favorable selection following the implementation of risk adjustment. It is also more than twice 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 and insurers' endogenous responses to that inattention.[8]

Beyond its bottom-line impact on public finances, the upcoding subsidy that we document is likely to distort Medicare beneficiaries' choice of health insurance away from TM. If the MA market is (imperfectly) competitive, then the upcoding subsidy effectively provides beneficiaries with a larger voucher for purchasing an MA plan than for purchasing TM. We show that increasing the competitiveness of the MA market can actually worsen this distortion and reduce net efficiency: Competition tips the incidence of the subsidy away from insurers and toward the consumer, but this shift in incidence distorts the consumer's choice between TM and MA.

Finally, our findings contribute to the growing policy literature on the broader welfare impacts of

[7]While this line does not matter for economic consequences, it may matter for insurer incentives to manipulate codes. Because CMS audits MA plan coding practices, the cost to an insurer of engaging in illegal "upcoding" is probably higher than the cost of legally changing coding practices to maximize profits.

[8]While, similar to Brown et al. (2014) and Ho, Hogan and Scott Morton (2014) we cannot perform a full welfare analysis of this coding difference, we note that the public spending implications are significant.

the MA program. Besides the benefits of expanding choice, one popular argument in favor of MA is that it might create important spillover effects for TM. Studies of physician and hospital behavior in response to the growth of managed care suggest the possibility of positive externalities in which the existence of managed care plans lowers costs for all local insurers (see for example, Baker, 1997; Glied and Zivin, 2002; Glazer and McGuire, 2002; Frank and Zeckhauser, 2000). Most recently, Baicker, Chernew and Robbins (2013) find that the expansion of MA resulted in lower hospital costs in TM. 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 distorted choices) of upcoding in MA.

The outline for the remainder of the paper is as follows. In Section 2, we provide a brief overview of how insurers can influence the diagnoses assigned to their enrollees, and we provide some suggestive evidence of relatively high coding intensity in MA. 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. Section 5 presents results, and Section 6 discusses several implications of our findings for policy and economic efficiency. Section 7 concludes.

2 Risk Adjustment and Endogenous Coding

2.1 Risk Adjustment

We begin by briefly describing the functioning of a risk-adjusted payment system in a regulated private insurance market. Plans receive a payment from a regulator for each individual they enroll, which supplements or replaces premiums paid by the enrollee. The net payment after risk adjustment R for enrolling individual i is equal to the individual's risk score, r_i, multiplied by some benchmark amount, ϕ, set by the regulator: $R_i = \phi \cdot r_i$.[9] The regulator distributes risk adjustment payments from a fund, or enforces transfers across plans.[10] The risk score itself is calculated by multiplying a vector of risk adjusters, $\boldsymbol{x}_i$, by a vector of risk adjustment coefficients, $\boldsymbol{\Lambda}$. Risk adjusted payments are therefore $R_i = \phi \cdot \boldsymbol{x}_i \boldsymbol{\Lambda}$.

In health insurance markets, risk adjusters $\boldsymbol{x}_i$ typically consist of a set of indicators for demo-

[9] 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.

[10] The fund can be financed via tax revenues or via fees assessed to health plans by the regulator.

graphic 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, these indicators are referred to as Hierarchical Condition Categories (HCCs). Below, we refer to $\boldsymbol{x}_i$ as conditions for simplicity. The coefficients $\boldsymbol{\Lambda}$ 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 $\boldsymbol{x}_i$ in some reference population. Coefficients are normalized 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 $\boldsymbol{x}_i$ 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-plan matches. We relax this assumption below and explore the implications.

2.2 Endogenous Coding

We define endogenous coding or "upcoding" to include any differences in coding practices across plans that would lead to two plans generating different risk scores for the same individual. Because coefficients $\boldsymbol{\Lambda}$ are set by the regulator and are fixed across insurers, upcoding can arise only from differences in the reporting of conditions. Formally, we relax the assumption that risk scores are exogenous to an enrollee's plan choice by allowing the reported conditions for individual i to vary by plan j. Below, we refer to the difference between the individual's risk score in plan j and her score in plan j', $r_i^j - r_i^{j'}$, as an individual-specific coding intensity difference. The difference in the risk adjusted payment for individual i between plans j and j' is

$$\Delta R_i = \phi \cdot (r_i^j - r_i^{j'}) = \phi \cdot \boldsymbol{\Lambda}(\boldsymbol{x}_i^j - \boldsymbol{x}_i^{j'}). \tag{1}$$

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

It is important to note that while financial incentives to code are often assumed to drive coding

intensity, calculating this difference in voucher size, ΔR_i, requires no assumptions about the source of coding differences. This is important because it implies that characterizing the public costs and the consumer choice distortions that arise from differential coding does not require defining some objectively correct level of coding. In our empirical setting below, it makes no difference whether physicians under TM pay too little attention to coding or whether MA insurers pay too much attention to coding. As long as the same individual would generate a higher risk score under an MA plan than under TM, the government payout is larger under MA, pushing consumers to the more generously subsidized MA market segment.[11] We discuss the choice distortion implied by this differential subsidy in more detail in Section 6.

2.3 Coding in Practice

In most markets with risk adjustment, regulators recognize the potential for upcoding and 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.[12] 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.

Figure 1 outlines the various mechanisms insurers employ to affect diagnosis coding, and in turn, risk scoring.[13] We leave out any mechanisms that involve illegal action on the part of insurers.[14] First,

[11] For example, suppose that there are two types of Diabetics: Severe and Moderate. Diabetics are distributed equally between these two groups and across TM and MA. The Severe (Moderate) Diabetics cost on average $30,000 ($10,000) per year to insurer. Now, suppose that the risk adjustment model has only one category for diabetes. In TM, only the Severe Diabetics get coded as having diabetes. Because the risk adjustment coefficients are estimated using data from TM, the coefficient for diabetes will be equivalent to $30,000. The incremental increase in the implicit TM voucher due to being coded as a diabetic is then $30,000 for Severe Diabetics and $0 for Moderate Diabetics. In MA, insurers ensure that coding is accurate and complete by coding both Moderate and Severe Diabetics as having diabetes. In this case, the incremental increase in the MA voucher due to being coded as a diabetic is $30,000 for both Severe and Moderate Diabetics. Thus, the MA voucher is equal to the TM voucher for the Severe Diabetics but it is $30,000 greater than the TM voucher for the Moderate Diabetics. This is true despite the fact that coding differences are due to greater vigilence with respect to coding accuracy on the part of the MA plans rather than to outright fraud.

[12] Traditionally, the diagnoses were included on the claim to provide justification for the service for which the provider was billing the insurer.

[13] 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.

[14] While fraud is likely a problem in health insurance markets, it is clear that coding differences can (and do) arise without any explicitly illegal actions on the part of the insurer.

and prior to any patient-provider interaction, insurers can structure contracts with physician groups such that the payment to the group is a fraction of the risk adjusted payment that the insurer itself receives from the regulator, directly passing coding incentives through to the groups. Insurers may also choose to selectively contract with providers who code more aggressively. Additionally, the insurer can influence coding during the medical exam itself by providing tools to the physician that pre-populate his notes with information on prior-year diagnoses for the patient. Since risk adjustment is based solely on the diagnoses from a single year, this increases the probability that diagnoses, once added, are retained through the next risk scoring period. Insurers also routinely provide training to the *physician's* billing staff on how to assign codes to ensure 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 request a change to the coding by the physician's billing staff, or directly alter the codes itself.[15]

Beyond these interventions with physicians and their staffs, insurers directly incentivize their enrollees to take actions that result in more intensive coding. Insurers may incentivize or require 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 and send a physician or nurse to the enrollee's home. The visit is both necessary in order to add the relevant, reimbursable diagnoses for the current plan year and relatively low cost. There is substantial anecdotal evidence for such behavior in Medicare Advantage,[16] and regulators have expressed serious concern that such visits primarily serve to inflate risk scores.[17]

None of these insurer activities take place in TM because providers under the traditional system are paid directly by the government, and, in the outpatient setting, these payments are based on procedures, not diagnoses.[18] In TM, diagnoses are instead used for the sole purpose of providing

[15] Insurers use a variety of software tools to scan medical records and determine for each individual the most lucrative combination of codes consistent with the medical record.

[16] See Center for Public Integrity 2014.

[17] 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."

[18] Under TM, hospitals are compensated for inpatient visits via the diagnosis-related groups (DRG) payment system, in which inpatient stays are reimbursed partially on the basis of inpatient diagnoses and partially on the basis of procedures.

justification for the services for which the providers are requesting reimbursement. This difference in incentive structure and between TM and MA naturally suggests that coding will be less intensive under TM, especially with respect to the codes that are relevant for payment in MA.

2.4 Suggestive Evidence

Before describing our identification strategy and main dataset, we briefly provide suggestive evidence of higher relative coding intensity in MA using a different dataset and strategy. In a series of summary statistics, we show how risk scores change when seniors turn 65 and join either MA or TM. This involves observing claims in an employer or commercial plan pre-65 and then claims for the same individual in MA or TM post-65.

To do this, we use a newly-available dataset including the universe of commercial health insurance claims in the state of Massachusetts from 2011 to 2012.[19] The dataset includes an individual identifier that allows us to follow people across health insurance plans, and it includes claims data from Medicare Advantage, Medigap, employer, and individual-market commercial insurers. We use a novel approach to identifying claims in TM, using data from Medigap plans. Because Medigap plans cover some fraction of almost every TM claim, they identify the claims of TM enrollees indirectly.[20]

We identify two groups of individuals 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.[21] We limit the sample to individuals with at least 6 months of data prior to and after joining MA/Medigap. Our final sample includes 4,724 Medigap enrollees and 1,347 MA enrollees. We use diagnoses from the claims data to generate risk scores for each individual in our sample just prior to and just after the age 65 threshold. Risk scores are calculated according

It is nonetheless plausible that overall coding intensity in TM 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 TM costs, claims, and diagnoses are established in the outpatient setting, in which physician reimbursement is depends on procedures, not diagnoses.

[19] Because the data used to construct the results in Figure 2 are different from the data used in our main analysis, we describe the construction of the statistics in more detail in Appendix A.2.

[20] The only claims that Medigap does not pay any part of are hospital *re*-admissions and lab claims (paid in full by TM). In our analysis we are assuming that these types of claims contain no relevant diagnoses that are not recorded on another claim for a particular beneficiary. For hospital readmissions, it is obviously unlikely that the new admission will include relevant diagnoses that didn't appear in a prior admission. For lab claims, it is unlikely that the lab claim itself includes a diagnosis that doesn't appear on the claim for the office visit corresponding to the lab test.

[21] In each of these groups, we remove everyone who is not continuously enrolled in MA/Medigap after their 65th birthday and everyone who is not continuously enrolled in a (non-Medicare) employer or commercial plan prior to their 65th birthday.

to the same Medicare Advantage HCC model regardless of the plan type in which the consumer is enrolled.[22]

In Figure 2 we present difference-in-differences summary statistics comparing the change in an individual's risk score at age 65 among consumers entering MA and consumers entering TM.[23] The top panels focus on the indicators for chronic conditions that are used to generate risk scores, and the bottom panel focuses on the risk scores themselves. Blue bars indicate individuals who choose to enroll in TM at age 65, and gray bars indicate individuals who choose to enroll in MA at age 65. Additional results from this sample are reported in Appendix Table A1.

All three panels of 2 suggest that individuals who choose to enroll in TM are sicker at age 64 than individuals who choose to enroll in MA: The TM enrollees have a higher probability of being coded with a chronic condition, are coded with more chronic conditions, and have higher risk scores prior to age 65, all as diagnosed by their non-Medicare, pre-65 insurance plan. This is consistent with many studies of advantageous selection into MA exploiting "switchers" (Newhouse et al., 2012; Brown et al., 2014). Consistent with the prevalence of chronic conditions increasing with age, both TM and MA enrollees appear sicker post-65 than they appear pre-65.

Suggestive evidence of upcoding is apparent in the differential change in diagnoses across the age 65 threshold for MA enrollees relative to the change for TM enrollees. All three panels of Figure 2 indicate that the diagnosed health status of MA enrollees appears to immediately and dramatically worsen at the time of enrollment, while the diagnosed health status of TM enrollees does not.[24] These summary statistics plotted in the figure imply that risk scores of MA enrollees increase 7.6% more than the increase in the risk scores of TM enrollees. In Appendix Table A1, we show these difference-in-differences comparisons are statistically significant and robust to including individual-level fixed effects.

These statistics suggest that coding intensity is higher in MA than in TM. However, it could be

[22] 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 enrolleed in MA/Medigap in July 2011, her pre period will consist of January-June 2011 and her post period will consist of January-June 2012. On the other hand, if an individual enrolled in MA/Medigap in February 2012, her pre period will consist of February-December 2011 and her post period will consist of February-December 2012.

[23] All statistics are based on risk scores that are adjusted for an individual's month of enrollment, gender, and the region of Massachusetts in which she resides.

[24] MA enrollees are coded for 0.15 additional chronic conditions and see a 6.4% increase in the probability of being coded with at least one chronic condition and a 0.067 point increase in their risk scores. TM enrollees, on the other hand, are only coded for 0.031 additional chronic conditions and only see a 0.5% increase in being coded with at least one chronic condition and a 0.023 point increase in their risk scores.

that the health status of MA enrollees has a steeper trajectory, as would be the case if individuals who expect their health status to worsen post-65 choose to enroll in MA rather than TM. This would present a violation of the parallel trends assumption needed for a difference-in-differences study around the age 65 threshold. Unfortunately, the short timespan of our dataset (2011-2012) precludes us from testing this assumption by examining pre-trends. Additionally, if MA has a negative causal effect on health, the observed changes in health status could be due to actual shifts in health rather than shifts in reported health.[25]

Even if the difference-in-differences estimator is unbiased, it is an estimate of the coding difference between 65 year-old MA enrollees and 65 year-old TM enrollees who purchased a Medigap plan. This is obviously a small subset of the MA and TM populations, and the result may not hold for the full Medicare population. Nonetheless, this analysis provides compelling suggestive evidence of higher coding intensity in MA than in TM. In the remainder of the paper, we develop and implement an identification strategy that deals with each of these issues and produces an unbiased estimate of the average coding difference between TM and MA across the entire Medicare population.

3 Identifying Upcoding in Selection Markets

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

Regulators have attempted to identify upcoding by observing how individual risk scores change when consumers switch from one plan to another (CMS, 2010).[26] The critical assumption behind this strategy is that health plan choice is exogenous to health trajectory. Because consumers are more likely to switch health plans in response to a health shock (Ho, Hogan and Scott Morton, 2014), this assumption is unlikely to hold in practice. An additional limitation of the switcher strategy is that it cannot be used to estimate a treatment effect among new enrollees. This shortcoming is non-

[25] However, it seems unlikely that any differences in practice patterns across TM and MA would present themselves so rapidly, especially given that our measures of health status are all based on the presence of chronic conditions.

[26] Switcher analysis has most prominently been used to examine selection in the Medicare Advantage market. See Newhouse et al. (2012) and Brown et al. (2014) for recent examples.

trivial in our setting given that most consumers choose between Traditional Medicare and Medicare Advantage upon becoming eligible for Medicare at age 65, and their choices exhibit a substantial amount of inertia (Sinaiko, Afendulis and Frank, 2013).

3.1 Graphical Intuition

Our solution to the identification problem is to focus on market-level risk. Assume that within a large geographic market, the total population distribution of *actual* health conditions is stationary. Market-level *reported* risk scores could nonetheless change if market shares shift between plans with higher and lower coding intensity.

Figure 3 provides the graphical intuition for this idea for the simple case in which selection generates average risk score curves that are linear in market shares. We depict two plans, or market segments, labeled A and B. These are intended to align with TM and MA, respectively. All consumers choose either A or B. Segment B 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.[27] The top panel shows three curves: the average risk in A ($\bar{r}^A$), the average risk in B ($\bar{r}^B$), and the average risk of all enrollees in the market ($\bar{r}$).

In the top panel of Figure 3 we plot the baseline case of no upcoding. The market share of B, denoted by θ^B, is increasing along the horizontal axis. Average risk in B is low at low levels of θ^B because the few beneficiaries selecting into B are the lowest risk. As long as there is no coding difference between A and B, the market-level risk averaged over enrollees in both plans, $\bar{r}$, is constant in θ^B. This is because reshuffling enrollees across plans within a market doesn't affect the market-level distribution of underlying health conditions.

The bottom panel of Figure 3 incorporates differential coding: For any individual, Segment B is assumed to assign a risk score higher than that assigned under A by some constant upcoding factor. For reference, the dashed line in the figure represents the counterfactual average risk that Segment B enrollees would have been assigned under Segment A coding practices, $\bar{r}^B_A$. The key insight is that here, unlike the top panel, the slope of market level risk $\bar{r}$ with respect to penetration θ^B is non-zero. Intuitively, $\left(\frac{\partial \bar{r}}{\partial \theta}\right)$ reflects upcoding because the marginal consumer switching plans from A to

[27] We ignore uncompensated selection, since our goal here is to distinguish between differences in the risk score due to coding and differences in the risk score due to compensated selection. "Compensated" here implies that the selection has no effect on the net cost curve that the insurer faces.

B increases θ^B and simultaneously increases the average reported risk in the market by moving to a plan that assigns her a higher score. While the bottom panel of Figure 3 depicts the empirically relevant case in which the advantageously selected segment is more intensely coded, we show next that the same intuition applies regardless of the presence or direction of selection.[28]

3.2 Model

We now generalize the graphical analysis above 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 between plans.

Continue to assume two plans or market segments, labeled A and B, from which all consumers choose. As in Section 2 let $\Delta_i \equiv r_i^B - r_i^A$ represent the difference between the risk score individual i would generate in plan B compared to A. Let θ^B denote the market share of B, and let $\mathbb{1}[B_i(\theta)]$ represent the indicator function for choosing B, which expresses, for any level of penetration θ^B, the plan choice of i. Then, the average risk score in the market can be expressed as

$$\bar{r} = \frac{1}{N} \sum \left(r_i^A + \mathbb{1}[B_i(\theta)](r_i^B - r_i^A) \right), \tag{2}$$

and the average counterfactual difference in coding, $\overline{\Delta}_i$, taken over all individuals in the market regardless of plan choice, can be expressed as $\frac{1}{N}\sum(r_i^B - r_i^A)$.

The top panel of Figure 3 showed that for the simple case in which selection generates linear plan-specific risk curves, homogenous coding across insurers implies that the market level average risk is constant in θ^B. To see that this holds generally, note that if coding is identical in Segments A and B, then $r_i^B = r_i^A$ for every enrollee, implying $\bar{r} = \frac{1}{N} \sum \left(r_i^A \right)$ and $\frac{\partial \bar{r}}{\partial \theta} = 0$.

The bottom panel of Figure 3 suggested that if coding differs between plans ($r_i^B \neq r_i^A$), then $\frac{\partial \bar{r}}{\partial \theta} \neq 0$. Under the assumption that an individual's risk score r_i^B is comprised of a plan-independent individual risk component $\hat{r}_i$ plus a plan-specific coding factor α_j, the slope of the market average risk curve $\bar{r}$ exactly pins down the coding difference between A and B:

$$\frac{\partial \bar{r}}{\partial \theta} = \alpha_B - \alpha_A. \tag{3}$$

[28] For illustration, in Appendix Figure A1, we depict a case in which coding differences exist absent any selection on the risk score, a case in which the adversely selected plan is more intensely coded, and a case in which selection is both nonlinear and non-monotonic.

Equation 3 implies that in the typical risk adjustment scheme in which risk scores are normed to one, if plan B generates risk scores 10% of the mean higher than what A would generate for the same consumers ($\alpha_B - \alpha_A = 0.10 \ \forall \ i$), then the slope of the market level risk curve with respect to B's penetration would be 0.10. The proof follows directly from the additive separability of the individual and plan-specific components of the risk score.[29] With additive separability, Equation 3 holds for any distribution of risks and for any form of selection across plans. In Appendix A.3, we show that Equation 3 also holds under the weaker assumption that any heterogeneity in coding at the individual $\times$ plan level is orthogonal to θ^B.

In the empirical exercise below, we assume that coding differences can be represented with the additive form ($r_i^j = \hat{r}_i + \alpha_j$). While our aggregate, market-level data cannot be used to identify heterogeneity in coding effects across individual enrollees, we discuss the implications of relaxing this assumption in Section 6. Briefly, if plans upcode different types of enrollees differently, and if this heterogeneity in coding is systematically linked to B's share of the market, θ^B, then the slope parameters we estimate are equal to the coding intensity difference for the marginal, rather than the average, enrollee. In this case, our estimates will represent local approximations to the mean coding difference.

4 Setting and Empirical Framework

We apply the identification insights from Section 3 to examine coding differences between Medicare Advantage plans and the Traditional Medicare program. We begin with an overview of the institutional features of payments to private plans in MA. Then we describe our data and discuss our identifying variation and empirical framework in detail.

4.1 Medicare Advantage Payments

Individuals who are eligible for Medicare can choose between the TM system administered by the federal government or coverage through a private plan chosen in the MA market. MA plans are attractive to Medicare beneficiaries because compared to the traditional system they offer more com-

[29] Proof: $\frac{\partial \bar{r}}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{1}{N} \sum (\hat{r}_i + \alpha_A + \mathbb{1}[B_i(\theta)](\alpha_B - \alpha_A)) = (\alpha_B - \alpha_A) \cdot \frac{\partial}{\partial \theta} \frac{1}{N} \sum \mathbb{1}[B_i(\theta)] = \alpha_B - \alpha_A$, which makes no assumption on the distribution of $\hat{r}_i$ or on joint distribution of risks and preferences that generate the selection curves $\bar{r}^A(\theta)$ and $\bar{r}^B(\theta)$.

prehensive financial coverage, such as lower deductibles and coinsurance rates, as well as additional benefits, such as dental care and vision care. The tradeoff faced by beneficiaries in choosing an MA plan is that most are managed care plans, which restrict enrollees to a particular network of doctors and may impose referral requirements and other mechanisms to limit access to specialists.

The regulator, the Centers for Medicare and Medicaid Services (CMS), makes monthly capitation payments to MA plans for each beneficiary enrolled. 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 $(\phi_c \cdot r_i^j)$ is a function of the benchmark rate ϕ_c, which varies across counties c, and a person-specific adjustment determined by an individual's risk score r_i^j, where i indexes consumers and j indexes insurers. Historically, county benchmarks have been set to capture the cost of enrolling the "national average beneficiary" in the Traditional Medicare program in the county, though Congress has made many *ad-hoc* adjustments over time.[30] CMS sets risk adjustment coefficients nationally using claims data from TM.

4.2 Data

Estimating the slope $\frac{\partial \bar{r}}{\partial \theta^{MA}}$ from Figure 3 requires observing market-level risk scores at varying levels of MA penetration. We obtained county-level averages of MA, TM, and total market risk scores from CMS for 2006 through 2011.[31] MA enrollment is defined as enrollment in any MA plan type, including managed care plans like HMOs and PPOs, Private Fee For Service, employer MA plans, and Special Needs Plans that serve Medicare-Medicaid dual eligibles.[32] Average risk scores within the MA and TM market segments are weighted by the fraction of the year each beneficiary was enrolled in the segment. We define MA penetration as the fraction of all beneficiary-months within a county spent in an MA plan during a given year. For most of our analysis, we collapse all MA plans together, and consider the markets as divided between the MA and TM segments, though we also analyze the partial effects of increased penetration among various subsets of MA plan types.

We supplement these county-level aggregates with administrative data on demographics for the

[30] In practice, benchmarks can vary by plan. See Appendix A.1 for full details.
[31] Similar data are unavailable prior to 2006.
[32] We exclude only enrollees in the Program of All-inclusive Care for the Elderly (PACE) plans.

universe of Medicare enrollees from the Medicare Master Beneficiary Summary File (MBSF) for 2006-2011. These data allow us construct county-level averages of the demographic component of risk scores, which we use below in a falsification test.[33]

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 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 TM, implying that MA selects healthier enrollees. We show below that MA enrollees are even healthier than they appear based on risk scores due to higher coding intensity in MA. To ease interpretation of the coefficient estimates below, in all regressions we scale the risk score variable so that the overall market-level average across counties equals exactly one. Table 1 also shows the dramatic increase in MA penetration over our sample period, which comprises our identifying variation.

4.3 Identifying Variation

4.3.1 MA Penetration Changes

We exploit the large and geographically heterogenous 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 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 4, 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 in almost all cases positive, though the size of these changes varied widely. Figure 5 shows that this MA penetration growth

[33] The demographic components (r_i^A) and diagnostic components (r_i^{Dx}) of individual risk scores are additively separable, which implies that the county averages of these are also additively separable: $\bar{r}_i^j = \frac{1}{n}\sum_{i \in I_c}\left(r_i^A + r_i^{Dx}\right) = \bar{r}_i^A + \bar{r}_i^{Dx}$. The j superscript is suppressed here for simplicity.

was not limited to certain regions or to urban areas. The figure shades each county according to its quantile of penetration changes.

Our identification strategy relies on year-to-year variation in penetration within geographic markets to trace out the slope of the market average risk curve, $\frac{\partial \overline{r}}{\partial \theta^{MA}}$. The identifying assumption is that these changes in MA enrollment are not correlated with changes in actual underlying population health. In particular, in the county fixed effects models we estimate below, this implies that year-to-year growth in MA enrollment in the county did not track year-to-year variation in the actual population-level health of the county. The assumption is plausible, given that county population health, reflected in the incidence of chronic conditions 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.[34] Further, we can test the assumption of no correlated underlying health trends with respect to a variety of independently observable demographic, morbidity, and mortality outcomes at the county level.

In our preferred specification, we use all of the within-market, over-time variation in MA penetration to identify the parameter $\frac{\partial \overline{r}}{\partial \theta^{MA}}$. As a robustness check, we also estimate versions of the main regressions that isolate different sources of year-to-year variation in MA enrollment. In particular, since our identification strategy is based on the expansion of MA following the 2006 introduction of MA-Part D plans, we control for changes in the MA market segment size arising from enrollment in *non*-Part D MA plans. This identifies estimates using only the penetration growth directly attributable to growth in the MA-Part D market, addressing concerns that other sources of MA penetration changes could be confounding estimates. 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 not offering a Part D drug benefit. This approach addresses the potential concern that county $\times$ years in which MA-Part D expanded more rapidly were differentially trending in underlying population health.

[34] We offer the following additional arguments for the plausibility of this identifying assumption. On the supply side, the assumption implies that insurers don't selectively enter counties or alter plan benefits based on year-to-year changes in the average health of the county. This seems sensible, given that the dramatic penetration growth over our period appears to be driven by regulatory changes to Medicare embodied in the Medicare Modernization Act of 2003. We would spuriously estimate upcoding effects in MA only if insurers expanded market share by lowering prices or increasing benefits in places where the population was simultaneously becoming sicker or older. In terms of consumer choice, our assumption implies that individuals' demand for MA does not increase as the average health in the county declines. This also seems plausible, as the literature suggests that if MA enrollees differ from TM enrollees, MA enrollees are healthier, not sicker (Brown et al. (2014), Newhouse et al. (2012)).

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 only be apparent with a lag relative to penetration changes.

We illustrate the timing in Figure 6. The individual's risk score that is used for payment throughout the calendar year $t+1$ is based on diagnoses from calendar year t. This implies, for example, that if an individual moves to MA from TM, the risk score for her entire first year in MA will be based on diagnoses she received while in TM 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 new Medicare eligibles 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 6, 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 a *2-year* lag. We exploit these timing features below.

4.4 Econometric Framework

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

$$\bar{r}_{sct} = \gamma_c + \gamma_t + \sum_{\tau \in T} \beta_\tau \cdot \theta^{MA}_{sc\tau} + f(X_{sct}) + \epsilon_{sct}, \tag{4}$$

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

The coefficients of interest are β_{t-1} and β_{t-2} because of the institutional feature described above in which risk scores are calculated based on the prior full year's medical history, so that upcoding could plausibly affect risk scores only after the first year of MA enrollment for prior TM enrollees and after the second year of MA enrollment for newly-eligible beneficiaries. Because of the short panel nature of the data—our data start in 2006 and end in 2011—in our main specification, we estimate β for only a single lag. Later, we report alternative specifications that include a second lag, though these necessarily decrease the sample size, limiting statistical precision. A positive coefficient on lagged penetration indicates more intensive coding in MA relative to TM. Under our additive coding intensity assumption, $\beta_{t-1} + \beta_{t-2}$ is exactly equal to $\alpha_{MA} - \alpha_{TM}$, the difference between the risk scores that the same individual would generate under the two systems.

We include the placebo regressor $\theta^{MA}_{sc,\tau=t}$ in all specifications. Because upcoding can plausibly affect market-level risk only with a lag, the contemporaneous effect of penetration changes on market-level risk reflected in β_t should be zero. The coefficient on the placebo reveals any source of contemporaneous correlation between MA penetration and unobservable determinants of county risk that could contaminate our results. Similar placebo tests can be performed for leads of penetration $\left(\theta^{MA}_{sc\,\tau>t}\right)$, again subject to the caveat of reducing the panel length.

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

5 Results

In this section, we examine whether reported average risk at the local market level increases with Medicare Advantage penetration.[35] We begin by presenting the results that include all MA plan types. After reporting on a series of falsification and placebo tests in support of our identifying assumption, we examine how upcoding effects vary according to the level of integration between the insurer and physician.

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 coefficient indicates that the total 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 Traditional Medicare. In column 2 we add linear state time trends, and in column 3 we add time-varying controls for county demographics.[36] Across specifications, the coefficient on lagged MA penetration is stable.

To put the size of these coding effects in context, an increase in market-level risk of 0.07 would imply that starting from a perfectly healthy county population, 6% of all people became paraplegic, 15% of all people contracted HIV, or 58% became diabetics. If, contrary to our identifying assumption, these estimates were simply capturing spurious correlation between actual changes in underlying health conditions in the local market and changes to MA penetration, it would require large negative shocks to market-level average health that closely tracked enrollment changes. More precisely, for correlation between health shocks and enrollment shocks to explain the results in Table 2, it would also require that the process determining enrollment changes (contract design by the firm and enrollment decisions by consumers) would have to precede the health shocks by one year, since plan characteristics are set and enrollment decisions are made in the fall of prior plan year.[37]

[35] In Appendix A.4 we perform an analogous exercise examining how the within-TM and within-MA average risk scores in a market vary with 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).

[36] These controls consist of 19 variables capturing the fraction of Medicare beneficiaries in the county-year in 5 year age bins from 0 to 85.

[37] This timing is unrelated to the lagged process by which risk scores can be influenced, described in Figure 6.

While these effects are large, they are not inconsistent with widely held beliefs about coding in MA. Since 2010, CMS has applied a 3.41% deflation factor to MA risk scores when determining payments to private plans in the MA program, under the assumption that private plans code the same patients more intensively. The Government Accountability Office has expressed concerns that coding differences between MA and TM are likely much higher, in the range of 5% to 7% (Government Accountability Office, 2013). However, neither agency—nor any other study—has been able to provide econometrically identified estimates of this coding difference.

As a robustness check, we isolate the component of our identifying variation that arises explicitly from the expansion of MA-Part D plans, which combined MA with the new prescription drug benefit introduced in 2006. To do so, in Appendix Table A3 we control for changes in MA penetration in any plan type other than MA-Part D. These regressions are identified solely from changes in MA-Part D enrollment, which expanded rapidly in the few years following Part D's introduction. We also isolate the complement of this variation by alternatively controlling for changes in MA-Part D penetration. All results are closely consistent with Table 2, which uses all within-county across-time variation in penetration.

5.2 Placebo Tests

The coefficient estimates for contemporaneous MA penetration in Table 2, which are close to zero and insignificant across all specifications, support our placebo test. These coefficients imply that the health of the population was not drifting in a way that was spuriously correlated with changes in penetration. In principle, we could extend the placebo test of our main regressions by examining leads in addition to the contemporaneous effect. In practice, we are somewhat limited by our short panel, which becomes shorter as more leads or lags are included in the regression. Due to the length of time the program has existed, our data extend back only to 2006. The most recent data year available is 2011. Therefore, including two leads and one lag of penetration restricts our panel to just 2007 to 2009. Nonetheless, in columns 1 through 3 of Table 3, we repeat the main analysis with additional leads, under the intuition that significant coefficients on contemporaneous effects or leads would provide evidence of confounding trends.

Column headers in Table 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

(θ_t^{MA}, θ_{t+1}^{MA}, and θ_{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, supporting the parallel trends assumption. 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 an impulse change in true population health.

As discussed in the context of Figure 6, switchers from TM to MA carry forward their old risk scores for one plan-year, and newly-eligible consumers aging into Medicare and choosing MA won't have risk scores based on diagnoses assigned in MA until after two plan years.[38] Column 4, which includes a second lag provides evidence consistent with this. Each coefficient in the table represents an independent effect, so that point estimates on the first and second lag of penetration in column 4 indicate a cumulative upcoding effect of 9.2% after two years. Unfortunately, in the short panel, we are restricted from looking at effects with longer lags or leads with any precision. Nonetheless, we report on an extended set of leads and lags in Appendix Table A4.

5.3 Falsification Tests

In Tables 4 through 6 we conduct a series of falsification tests intended to uncover any correlation between changes in MA penetration and changes in other time-varying county characteristics. In particular, we focus on county demographics, mortality, and morbidity, since correlations between these characteristics and MA penetration could undermine the identifying assumption.

Table 4 replicates the specifications in columns 1 through 3 of Table 2, but with the demographic portion risk scores as the dependent variable. The demographic portion of the risk score is based only on age and gender, and unlike diagnoses is not manipulable by the insurer because CMS retrieves this information from Social Security data. The coefficients, which are near zero and insignificant in all specifications, show no impact of lagged penetration, consistent with the mechanism we describe in which enrollees are assigned more, or more severe, medical conditions.[39]

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

[39] An alternative valid interpretation of the results in Table 4 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.

In Table 5, we next test whether changes in MA penetration are correlated with measures of mortality and morbidity. Columns 1 through 3 show the relationship between changes in a county's mortality rate and changes in MA penetration. For morbidity, we use cancer incidence data from the Surveillance, Epidemiology, and End Results (SEER) Program by the National Cancer Institute. Columns 4 through 6 show the relationship between cancer incidence in a county-year and MA penetration. Cancer data is limited to the subset of counties monitored by SEER, which accounted for 27% of the US population in 2011, and 25% of the population over 65. Coefficients on both contemporaneous and lagged MA penetration are consistently close to zero and statistically insignificant.

Finally, we test whether changes in MA penetration are correlated with changes in the county's Medicare enrollee age distribution. In Table 6, the dependent variables measure the fraction of a county's Medicare population within each specified age range. The estimates show no consistent evidence of a systematic relationship between MA penetration and the Medicare enrollee age distribution. For completeness, we repeat this analysis using the county's population age distribution unconditional on Medicare status in Appendix Table A5. In sum, each falsification test supports our identifying assumption of no correlation between MA penetration and actual underlying population health or demographics, conditional on our controls.

5.4 Heterogeneity by Insurer and Physician Integration

We have so far treated all MA plans symmetrically, but the power insurers can exert over provider coding patterns is likely to vary depending on the level of integration between the insurer and the physician groups with whom the insurer contracts. We consider plans that are physician owned, use physician networks, or follow managed care models (i.e. HMOs and PPOs) to exhibit greater insurer/provider integration. These plans are likely to have more tools available for influencing provider coding patterns. For example, these plans may be able to pay their physicians partly or wholly as a function of the risk score that physicians' diagnoses generate. Additionally, these plans may be able to use the threat of removal from their networks to increase their bargaining power with physicians when negotiating the terms of their contracts, which can include incentives for coding intensity. Integration, therefore, could affect a plan's capacity to code intensely.

While we cannot observe integration directly, we proxy for integration with plan type. HMOs may be the most likely to exhibit integration, followed by PPOs. Private fee for service (PFFS) plans

are fundamentally different. During most of our sample period PFFS plans did not have networks of providers. Instead, they reimbursed Medicare providers on the basis of procedure codes (not diagnoses) at standard Medicare rates. Thus, PFFS plans had access to only a subset of the tools available to managed care plans for influencing the recording of diagnoses within the physician's practice. In particular, PFFS insurers could not arrange a contract with providers that directly rewarded intensive coding, nor would PFFS insurers be likely to train a physician's billing staff on coding. PFFS 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 TM.

In Table 7 we separate out the effect of penetration increases in HMO, PPO, and PFFS plans. We also separately control for employer-sponsored MA plans.[40] Employer plans are less strongly incentivized to code intensively due to the fact that in these plans financial insurance is provided by the employer with claims and diagnosis-reporting typically administered by third-parties. This adds a layer of principal-agent problems into the coding incentive structure.[41]

As in the main analysis, the coefficients of interest in Table 7 are on lagged penetration, while contemporaneous penetration coefficients comprise our standard placebo test. The table shows that the strongest coding intensity is associated with managed care plans, 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 TM. PPO coding intensity is around 7% higher than TM. Risk scores in PFFS and employer plans, while intensely coded relative to TM, exhibit relatively smaller effects.

6 Discussion

6.1 Public Finance Impacts

The implicit subsidy to Medicare Advantage due to differences in coding intensity is equal to the upcoding factor multiplied by the county benchmark rate ($\phi(\alpha_{MA} - \alpha_{TM})$).[42] In 2014, the average

[40] These regressions also control for penetration by the remaining plan types. The MA program includes a variety of 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.

[41] A typical MA insurer only needs to exert its influence on providers to increase coding intensity. However, an employer has to influence the third party administrator to influence providers, likely a more difficult task.

[42] This statement relies on the assumption that risk scores differ across MA and TM according to additive plan-specific coding factors (or the assumption that any heterogeneity in upcoding factors is orthogonal to MA penetration). If this assumption is relaxed, the parameter we estimate is the upcoding factor of the marginal enrollee. Thus, our back-of-the-

annual value of ϕ was about \$10,000. Given our estimate of $\alpha_{MA} - \alpha_{TM} = 0.07$ for MA overall, this implies a subsidy of about \$700 per MA enrollee in 2014, or a total potential excess subsidy of about \$11.4 billion, absent any coding inflation factor.[43] In 2010, CMS put in place a 3.41% inflation factor. Our results suggest this is both too small and fails to account for large coding differences across insurance contract types. While PFFS plans and employer MA plans differ in coding intensity by 4-5% relative to TM, PPOs inflate risk scores by 6-7% and HMOs, which comprise the largest category of MA plans, inflate risk scores by 10%. This translates to an incremental implicit subsidy of about \$1,000 per MA enrollee annually to HMOs.

These costs of differential coding are substantial. Brown et al. (2014) study the uncompensated advantageous selection into MA caused by the implementation of risk adjustment in 2006. They show 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 TM because MA plans attracted relatively low cost enrollees conditional on their risk scores. Our results imply that the Brown et al. (2014) estimate, while large and important in itself, dramatically understates the problem of implicit overpayments to MA plans arising from risk adjustment because that study doesn't consider heterogenous coding.

Recent changes to Medicare that are intended to improve quality of care in TM are likely to have the unintended consequence of encouraging higher coding intensity by TM providers. Interestingly, this might have the benefit of reducing excess payments to MA plans. Newly established Accountable Care Organizations (ACOs) under TM are intended to incentivize cost savings by making providers the residual claimants on a portion of the healthcare savings they generate for their patients. In the Medicare Shared Savings Program, the most popular ACO program, regulators compensate TM physician and hospital groups when patients under their care use relatively fewer resources relative to their *risk adjusted* health state, potentially incentivizing more intense coding in order to maximize payments under this metric.[44]

Higher coding intensity in TM relative to the status quo could actually reduce the budgetary

envelope calculations of the implicit subsidy will be accurate if (i) the additive plan-specific coding factor assumption holds, (ii) any heterogeneity in upcoding factors is orthogonal to MA penetration, or (iii) the average upcoding factor among the marginal group of enrollees in our analysis approximates the average upcoding factor among all MA enrollees in 2014. We assume that at least one of these three conditions holds.

[43] Based on September 2014 enrollment of 16,347,808 beneficiaries in MA, reported in Monthly SCC Enrollment Files provided by CMS.

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

impacts of intensive coding in MA. Because the relevant diagnoses for ACO reimbursements overlap with those in the MA risk adjustment scheme, coding incentives would be more closely aligned between MA and TM. From Eq. 1, more intense coding under TM would reduce the subsidy wedge between TM and MA by the amount $\phi(\alpha_{TM}^{\text{new}} - \alpha_{TM}^{\text{old}})$.[45] Such a possibility illustrates our claim that most public finance consequences of "upcoding" are completely determined by relative differences in coding intensity. The notion of an objectively correct level of coding has little economic content.

6.2 Welfare Consequences

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

While this paper cannot systematically address the possibility that intensive coding itself represents a good, we expect such impacts to be relatively small for several reasons. For one, insurers primarily intervene with physicians' coding and billing staff rather than with the physicians themselves. Insurers influence the coding of diagnoses in *insurance claims*, to which the physician, who gleans information instead from patient charts, may have no exposure. Second, coding aimed at maximizing payment and coding aimed at improving quality and continuity of care are likely to be different. One large MA insurer explained to us that the type of documentation physicians desire for clinical reasons is often at odds with what insurers desire for risk-scoring.[46] Finally, the regulatory agencies that oversee risk adjustment programs express serious doubt about the clinical value of insurer activities surrounding coding, some of which lack any connection to the actual provision of healthcare. For example, in a 2014 statement, CMS explained its view that home health visits and risk assessments "are typically conducted by healthcare professionals who are contracted by the vendor and are not part of the plan's contracted provider network, i.e., are not the beneficiaries' primary care providers."

If intensive coding doesn't meaningfully affect consumer well-being by influencing real care pro-

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

[46] Source: Our interview with an anonymous large managed care MA insurer on April 4, 2014.

vision, then the primary welfare consequences of coding intensity differences are the distortions these differences introduce into health plan choice. In our empirical setting, seniors are presented with an implicit voucher to purchase an MA plan that is significantly larger than the value of their implicit TM voucher. Similarly, within the MA market segment choices are distorted towards the plans and plan types with higher coding intensity, namely HMO plans.

A full welfare analysis of the TM/MA choice margin is beyond the scope of this paper, though distorting consumer choices towards MA via this subsidy could be efficient if (i) frictions caused lower than optimal enrollment in MA based on private benefits,[47] or (ii) MA enrollment has important external benefits, such as the cost control spillovers documented by Baicker, Chernew and Robbins (2013). In any case, understanding the size of the implicit coding subsidy that we calculate here would be key to any full welfare analysis.

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

In terms of consumer choice, stronger MA competition may paradoxically exacerbate the sorting distortion. If competition is stronger, then insurers will pass-through the coding subsidy to consumers in the form of lower premiums or additional benefits (Cabral, Geruso and Mahoney, 2014).[48] While this increases consumer surplus at the cost of producer surplus, it may also decrease net efficiency by distorting choices. On the other hand, if competition is weaker, insurers will pass-through a smaller fraction of the overpayment, resulting in a smaller distortion to seniors' choice between TM and MA. Thus, competition results in a tradeoff between the incidence of the subsidy and net efficiency: If competition is strong, the implicit coding subsidy will result in additional consumer surplus but distorted choices; if competition is weak, the subsidy will result in additional producer surplus but a smaller choice distortion.

[47] For example, inertia to remain in Traditional Medicare.

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

In other regulated markets where private insurers don't compete against a public option, such as the ACA Exchanges, upcoding nonetheless remains a significant problem. In these markets, risk adjustment has no first order impact on public budgets, because a regulator simply enforces transfers from plans with lower average risk scores to plans with higher average risk scores.[49] In these settings, plans are incentivized to code intensively to maximize profits, trading off the incremental subsidy to intensive coding against its cost. If coding intensity has no direct impact on consumer welfare, then investment in coding practices implies a deadweight loss. Perhaps more importantly, if there is heterogeneity across plans in the cost of increasing coding intensity, then upcoding-induced choice distortions will persist even in fully competitive, fully private insurance markets. Our result that MA plans with higher levels of insurer-provider integration display higher coding intensity suggests that the choices of Exchange enrollees are likely to be inefficiently distorted toward these more integrated plans.

7 Conclusion

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

Risk adjustment addresses an important problem of asymmetric information in insurance markets. Therefore, in the second-best world in which adverse selection is an inherent feature of competitive insurance markets, the optimal payment mechanism may include some kind of risk adjustment despite the costs and distortions of manipulable coding that we document. Nonetheless, our study offers some insight into improvements in risk adjustment mechanism design: From the perspective of this paper, the risk adjustment literature focusing on the predictive content of risk scores is pursuing the wrong objective function. Glazer and McGuire (2000) show that to induce efficient health

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

plan benefit design, risk adjustment must focus on insurer incentives rather than predicting expected costs. Applied to our findings, this suggests that the (second-best) optimal payment policy may include risk adjustment, but with coefficients on risk adjusters that account for both predictiveness of costs and susceptibility to differential coding. 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 may be unobservable, fewer and coarser diagnosis groups might be a feasible alternative preferable to current risk adjustment systems. This is an issue of significant practical importance, given the large and growing role of risk adjustment in regulated insurance markets for Medicare, Medicaid, and Exchange plans.

References

Akerlof, George A. 1971. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics*, 84(3): 488–500.

Baicker, Katherine, Michael E. Chernew, and Jacob A. Robbins. 2013. "The spillover effects of Medicare managed care: Medicare Advantage and hospital utilization." *Journal of Health Economics*, 32(6): 1289 – 1300.

Baker, Laurence. 1997. "HMOs and Fee-For-Service Health Care Expenditures: Evidence from Medicare." *Journal of Health Economics*, 16: 453–482.

Brown, Jason, Mark Duggan, Ilyana Kuziemko, and William Woolston. 2014. "How does risk selection respond to risk adjustment? Evidence from the Medicare Advantage program." *American Economic Review*, 104(10): 3335–3364.

Buchmueller, Thomas, and John DiNardo. 2002. "Did Community Rating Induce an Adverse Selection Death Spiral? Evidence from New York, Pennsylvania, and Connecticut." *The American Economic Review*, 92(1): 280–294.

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*.

Cao, Zhun, and Thomas G McGuire. 2003. "Service-level selection by HMOs in Medicare." *Journal of Health Economics*, 22(6): 915–931.

Carey, Colleen. 2014. "Government Payments and Insurer Benefit Design in Medicare Part D."

Carlin, Caroline, and Robert Town. 2010. "Adverse selection: The dog that didn?t bite." University of Minnesota.

Chetty, Raj, John N. Friedmand, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104(9): 2593–2632.

CMS. 2010. "Announcement of Calendar Year (CY) 2010 Medicare Advantage Capitation Rates and Medicare Advantage and Part D Payment Policies." Centers for Medicare and Medicaid Services.

Dafny, Leemore S. 2005. "How Do Hospitals Respond to Price Changes?" *American Economic Review*, 1525–1547.

Einav, L, and A Finkelstein. 2010. "Selection in Insurance Markets: Theory and Empirics in Pictures." *Journal of Economic Perspectives*, 25(1): 115–138.

Einav, Liran, Amy Finkelstein, and Jonathan Levin. 2010. "Beyond Testing: Empirical Models of Insurance Markets." *Annual Review of Economics*, 2: 311–36.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen. 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices." *The Quarterly Journal of Economics*, 125(3): 877–921.

Frank, Richard G, and Zeckhauser. 2000. "Custom-made versus ready-to-wear treatments: Behavioral propensities in physicians' choices." *Journal of Health Economics*, 26(6): 1101–1127.

Frank, Richard G, Jacob Glazer, and Thomas G McGuire. 2000. "Measuring adverse selection in managed health care." *Journal of Health Economics*, 19(6): 829–854.

Geruso, Michael. 2012. "Selection in Employer Health Plans: Homogeneous Prices and Heterogeneous Preferences."

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 Adjustmnets." Accessed on Oct 8, 2013 at http://www.gao.gov/assets/660/651712.pdf.

Hackmann, Martin B, Jonathan T Kolstad, and Amanda Ellen Kowalski. 2013. "Adverse Selection and an Individual Mandate: When Theory Meets Practice." National Bureau of Economic Research, Inc.

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, Igal Hendel, and Michael D Whinston. 2013. "Equilibria in Health Exchanges: Adverse Selection vs. Reclassification Risk." National Bureau of Economic Research.

Ho, Kate, Joe Hogan, and Fiona Scott Morton. 2014. "The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program."

Layton, Timothy J. 2014. "Imperfect Risk Adjustment, Risk Preferences, and Sorting in Competitive Health Insurance Markets."

Lustig, Josh. 2011. "Measuring welfare losses from adverse selection and imperfect competition in privatized Medicare."

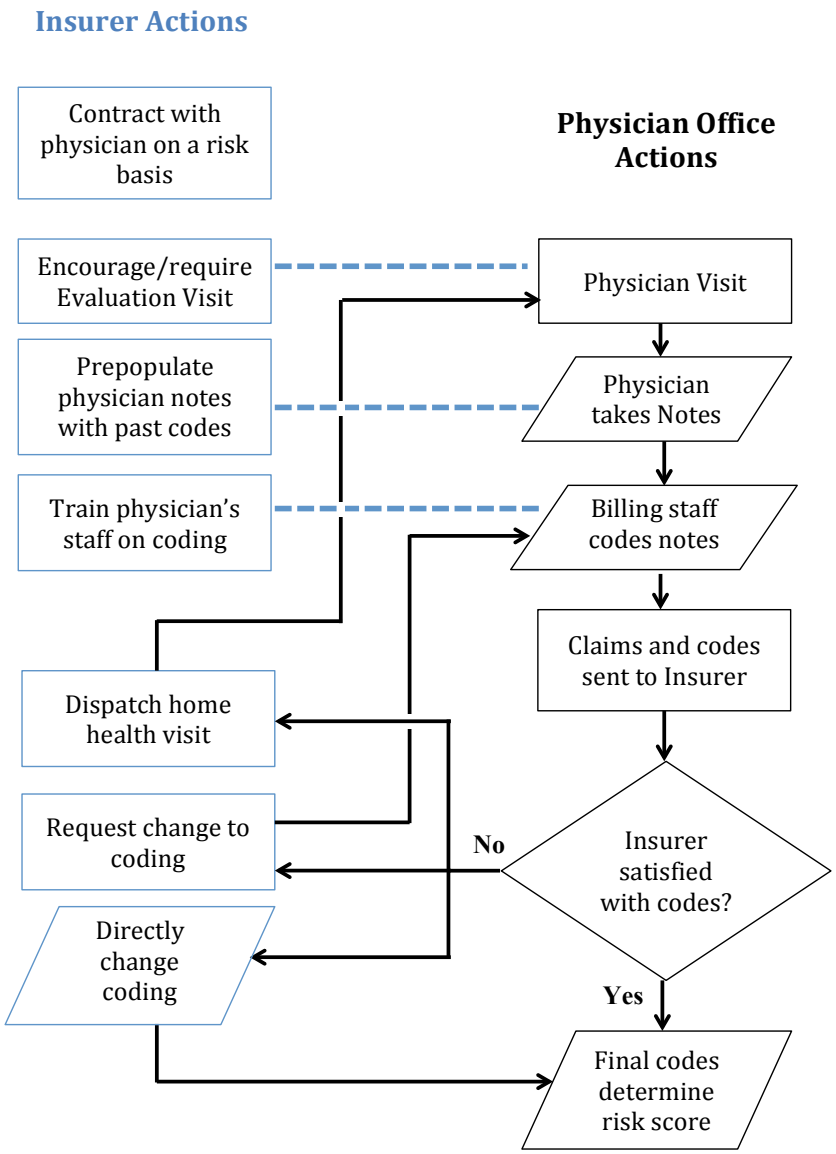
Newhouse, Joseph P, Mary Price, Jie Huang, J Michael McWilliams, and John Hsu. 2012. "Steps to reduce favorable risk selection in Medicare advantage largely succeeded, boding well for health insurance exchanges." *Health Affairs*, 31(12): 2618–2628.

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *The Quarterly Journal of Economics*, 90(4): 629–649.

Silverman, Elaine, and Jonathan Skinner. 2004. "Medicare upcoding and hospital ownership." *Journal of health economics*, 23(2): 369–389.

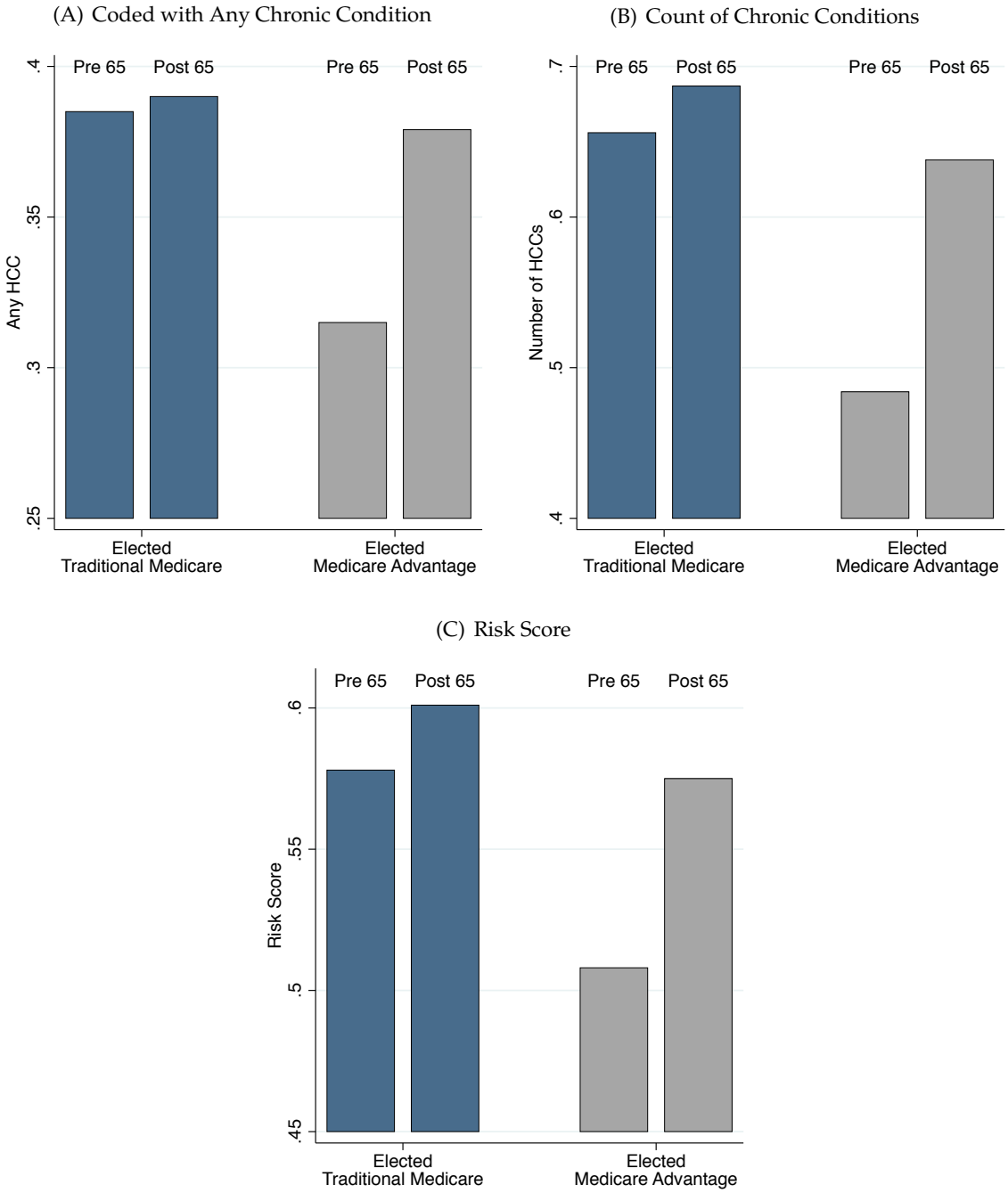
Sinaiko, Anna D, Christopher C Afendulis, and Richard G Frank. 2013. "Enrollment in Medicare Advantage Plans in Miami-Dade County Evidence of Status Quo Bias?" *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, 50(3): 202–215.

Figure 1: How Risk Scores are Influenced by Insurers



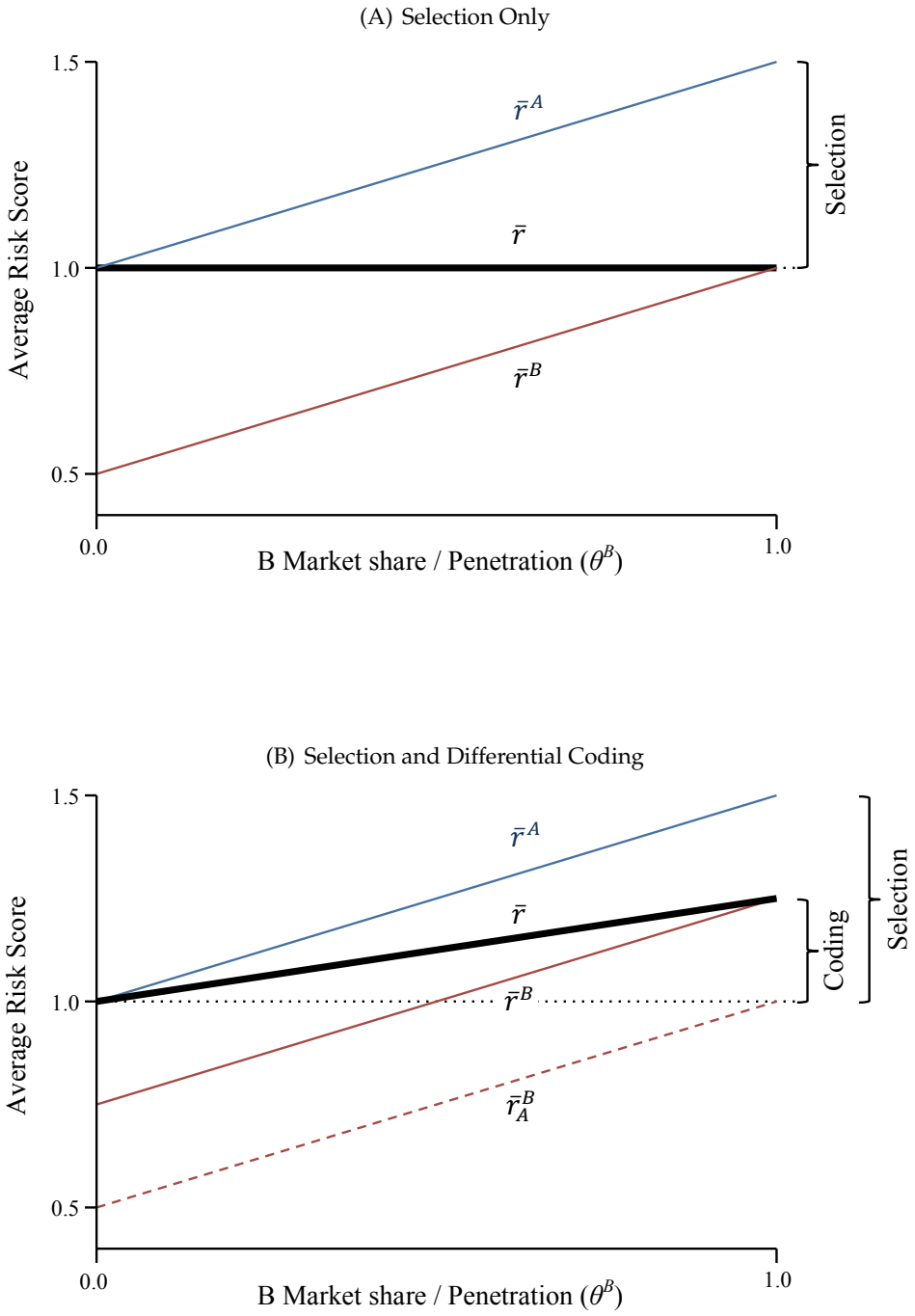
Note: The flowchart illustrates how diagnosis codes originate and how insurers can influence the process that generates them. Insurer actions are towards the left of the figure in blue boxes. Provider actions, including those of the provider's billing and coding staff are towards the right in black boxes.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

Figure 2: Coding Across the Age 65 Threshold for Beneficiaries Choosing TM and MA



Note: 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 TM. Blue bars indicate individuals who choose to enroll in TM at age 65, and gray bars indicate individuals who choose to enroll in MA at age 65. Sample includes 4,724 Medigap enrollees and 1,347 MA enrollees. We describe the construction of these statistics in more detail in Appendix A.2.

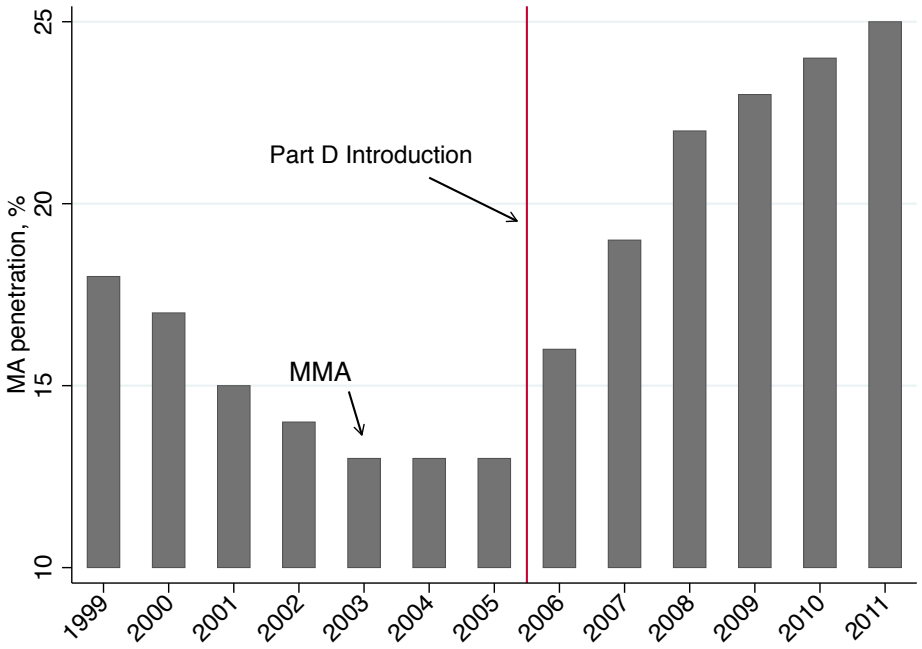
Figure 3: Identifying Coding Differences in Selection Markets



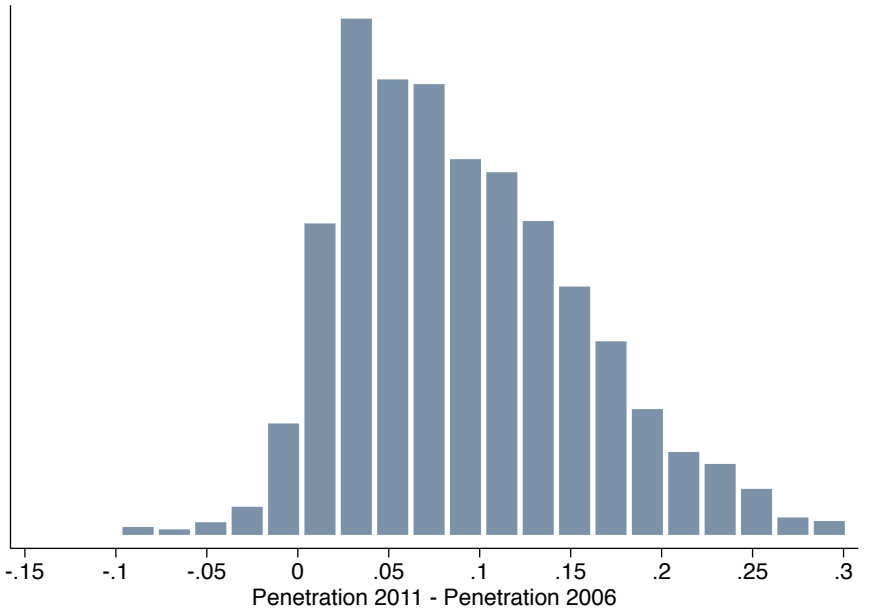
Note: This figure demonstrates how to separate coding differences from selection when true underlying risk is unobservable. The horizontal axis measures the market share of segment B, θ^B. The vertical axis measures the average risk score: Average risk in A is $\bar{r}^A$, average risk in B is $\bar{r}^B$,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

Figure 4: Growth in Medicare Advantage (MA) Penetration

(A) National MA Penetration Changes Following MMA Implementation

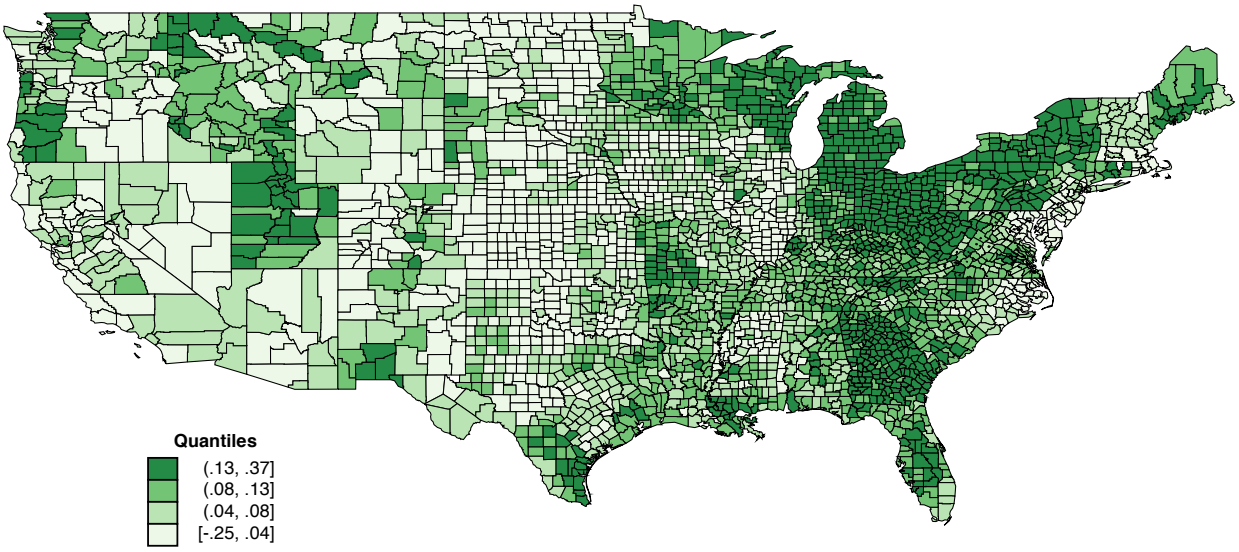


(B) Within-county MA Growth



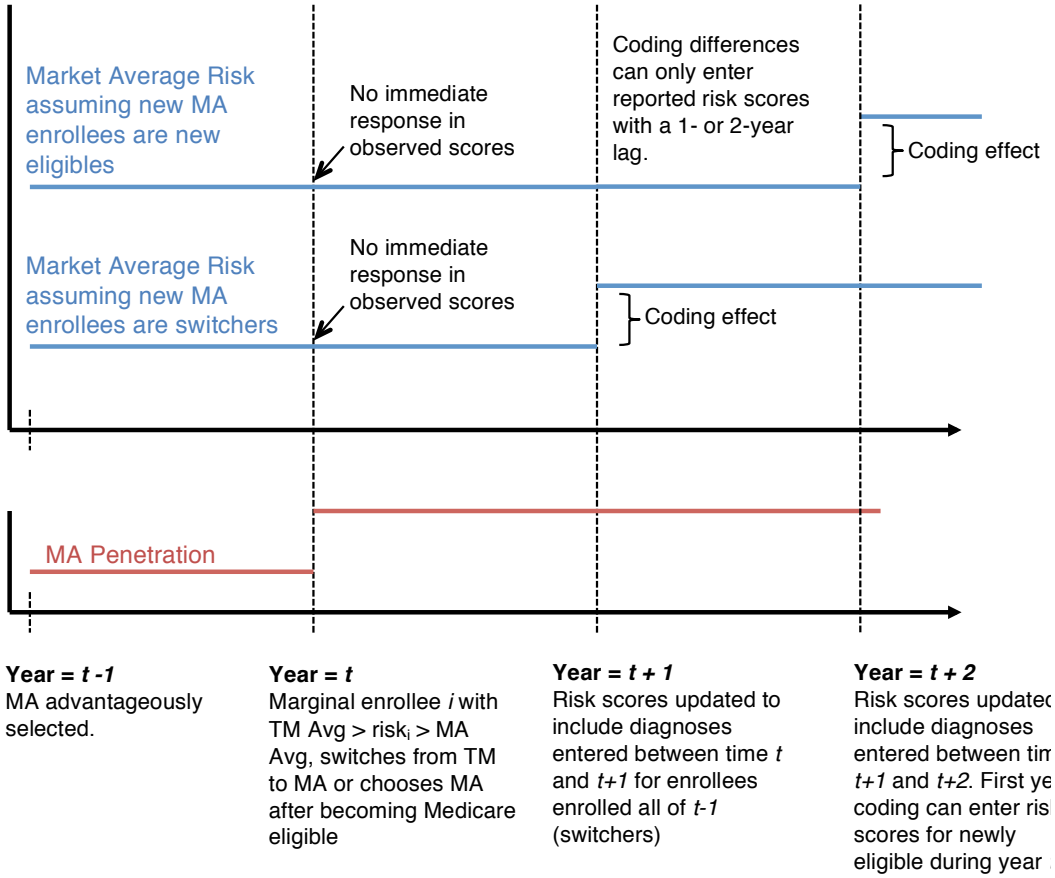
Note: The top panel displays national trends in MA penetration, where the unit of observation is the Medicare beneficiary. *Source, pre 2006:* Kaiser Family Foundation, 2013. The bottom panel displays a histogram of within-county changes in penetration from 2006 to 2011, using the main estimation sample.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

Figure 5: Geography of Growth in Medicare Advantage (MA), 2006 to 2011



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

Figure 6: Timing Illustration: Coding Effects Occur with a Lag in Medicare



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

Table 1: Summary Statistics

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

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

Table 2: Main Results

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

Note: Table shows coefficients on contemporaneous (t) and lagged ($t-1$) Medicare Advantage (MA) penetration. The dependent variable is the average risk score in the market (county). Because MA risk scores are calculated using diagnosis data from the prior plan year, coding differences can plausibly affect risk scores only with a lag, and coefficients on contemporaneous penetration serve as a placebo test. 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 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: Effects of Contemporaneous Penetration and Leads

	Dependent Variable: County-Level Average Risk Score			
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	0.032	
		(0.025)	(0.056)	
MA penetration t (placebo)	0.001	-0.021	-0.064	0.006
	(0.019)	(0.028)	(0.071)	(0.017)
MA penetration t-1	0.064**	0.076**	0.084**	0.041**
	(0.011)	(0.018)	(0.022)	(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-Year Demographics	X	X	X	X
Observations	15,640	12,512	9,384	12,512

Note: Table shows coefficients on future ($t+2$, $t+1$), contemporaneous (t), and lagged ($t-1$, $t-2$) Medicare Advantage (MA) penetration. The dependent variable is the average risk score in the market (county). Because MA risk scores are calculated using diagnosis data from the prior plan year, coding differences can plausibly affect risk scores only with a lag, and coefficients on contemporaneous and future penetration serve as a placebo test. Observations are county $\times$ years. 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. 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	0.001	0.001
	(0.002)	(0.002)	(0.002)
MA penetration t-1	0.001	0.000	-0.001
	(0.002)	(0.002)	(0.002)
Main Effects			
County FE	X	X	X
Year FE	X	X	X
Additional Controls			
State X Year Trend		X	X
County-Year Demographics			X
Mean of Dep. Var.	0.485	0.485	0.485
Observations	15,640	15,640	15,640

Note: Table shows coefficients on contemporaneous (t) and lagged ($t-1$) Medicare Advantage (MA) penetration. The dependent variable is the average *demographic* risk score in the market (county). Demographic risk scores are calculated by the authors using data from 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 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	-0.005	-0.005	-0.005
	(0.002)	(0.002)	(0.003)	(0.004)	(0.005)	(0.005)
MA penetration t-1	0.002	-0.002	-0.002	0.005	0.001	0.003
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.005)
Main Effects						
County FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Additional Controls						
State X Year Trend		X	X		X	X
County-Year Demographics			X			X
Mean of Dep. Var.	0.048	0.048	0.048	0.023	0.023	0.023
Observations	15,408	15,408	15,408	3,050	3,050	3,050

Note: Table shows coefficients on contemporaneous (t) and lagged ($t-1$) Medicare Advantage (MA) penetration. The dependent variable in columns 1 through 3 is the morality rate, derived from the National Center for Health Statistics, among county residents ≥ 65. The dependent variable in columns 4 through 6 is the cancer incidence rate among county residents ≥ 65. Cancer incidence data come from the Surveillance, Epidemiology, and End Results (SEER) Program by the National Cancer Institute. 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: Falsification Test: Effects on Medicare Age Distribution

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

Note: Table shows coefficients on contemporaneous (t) and lagged ($t-1$) Medicare Advantage (MA) penetration. The dependent variable in column 1 is the fraction of the Medicare population ≥ 65. The dependent variables in columns 2 through 6 are the fraction of the Medicare population in age bins conditional on age ≥ 65. 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 7: 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	0.035	0.026
		(0.033)	(0.036)	(0.035)
PPO penetration	t (placebo)	-0.006	0.011	0.012
		(0.033)	(0.038)	(0.037)
PFFS penetration	t (placebo)	-0.015	0.000	0.001
		(0.032)	(0.037)	(0.036)
Employer MA penetration	t (placebo)	0.002	-0.022+	-0.021
		(0.011)	(0.013)	(0.013)
HMO penetration	t-1	0.137**	0.098**	0.101**
		(0.027)	(0.028)	(0.028)
PPO penetration	t-1	0.115**	0.072*	0.068*
		(0.026)	(0.028)	(0.028)
PFFS penetration	t-1	0.048*	0.063*	0.058*
		(0.023)	(0.025)	(0.025)
Employer MA penetration	t-1	0.036**	0.041**	0.041**
		(0.010)	(0.012)	(0.012)
Main Effects				
County FE		X	X	X
Year FE		X	X	X
Additional Controls				
State X Year Trend			X	X
County-Year Demographics				X
Observations		15,640	15,640	15,640

Note: Table shows coefficients on contemporaneous (t) and lagged ($t-1$) Medicare Advantage (MA) penetration, disaggregated by contract type. The dependent variable is the average risk score in the market (county). Because MA risk scores are calculated using diagnosis data from the prior plan year, coding differences can plausibly affect risk scores only with a lag, and coefficients on contemporaneous penetration serve as a placebo test. Observations are county $\times$ years. 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. Other controls are as described in Table 2. Standard errors in parentheses are clustered at the county level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

APPENDIX

A.1 Background on MA Risk-Adjusted Payments

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

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

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

A.2 Data Used in Section 2

For the analysis in Section 2, we obtained data on Medicare Advantage and Medigap claims from the new Massachusetts All-Payer Dataset (Mass APCD) from the Massachusetts Center for Health Information and Analysis (CHIA). The Mass APCD includes the universe of health insurance claims for individuals receiving medical services in the state of Massachusetts. Payers, along with third-party claims administrators and pharmacy benefit managers, report all claims to the state of Massachusetts. These claims are then aggregated into a large, comprehensive dataset. To identify individuals, we use an individual ID created by the state generated using Social Security numbers. 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 1 month of their 65th birthday. We identify Medicare Advantage plans using an identifier provided by CHIA. We verify that the products are Medicare Advantage products by matching the names of the payers in the Mass APCD data to publicly available Medicare Advantage enrollment data provided by CMS and by observing the age distribution of enrollees in these plans, focusing on the portion of enrollees over age 65. The second group consists of all individuals enrolling in a product identified as a Medigap plan within 1 month of their 65th birthday. We also identify Medigap plans using an identifier provided by CHIA. Again, we verify that the products are Medigap products 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 most likely 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 the products we identified as MA and Medigap products. Because Medicaid data is excluded from our version of the Mass APCD, this results in a sample of individuals with continuous commercial coverage prior to their 65th birthday and continuous Medigap/MA coverage after their 65th birthday. To ensure that we have enough data to calculate risk scores, we restrict the sample to individuals with at least 6 months of data prior to joining Medigap/MA and 6 months of data after joining Medigap/MA. This effectively eliminates anyone joining Medigap/MA between January and June 2011 and between August and December 2012.

We retrieve the claims for each individual in the sample. We restrict the claims sample to claims from medical plans only, excluding prescription drug plans.[50] 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 those 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.[51]

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 program provided by CMS. The program maps each individuals diagnoses to a set of 70 chronic conditions. It then multiplies indicators for these chronic conditions by a set of weights estimated using claims data from Traditional Medicare. The product of the individual's chronic conditions indicators and the weights is the risk score. We also use the chronic conditions indicators to calculate each individual's number of chronic conditions and to generate an dummy variable indicating whether the individual has at least one chronic condition.

To produce the statistics reported in Section 2, we adjust the risk scores for gender and the region of Massachusetts in which the individual resides. We also adjust for cohort fixed effects to account for the fact that the risk scores are calculated using time periods of different lengths. We do the adjustment by running a regression of risk scores on these variables and extracting the residuals. The adjusted risk scores are equal to the residuals plus the estimated intercept from the regression. Using individual fixed effects produces similar results.

A.3 Linearity of Market-level Average Risk Curve

Throughout Section 3 we impose the assumption that coding differences can be represented by an additive factor. Under this assumption, the slope of the average risk curve in the market is exactly equal to the upcoding factor. It is true, however, that the weaker assumption that any individual-specific heterogeneity in the plan coding factor is orthogonal to Segment B's market share leads to the similar result: The slope of the average risk curve in the market is exactly equal to the average upcoding factor across individuals in the market.

To formalize this statement, let individual i's risk score in Plan A be equal to $r_i^A = \hat{r}_i + \alpha_A$, and let i's risk score in Plan B be equal to $r_i^B = \hat{r}_i + \alpha_B + \epsilon_{iB}$ where ϵ_{iB} has mean $\bar{\epsilon}$ and represents

[50] The CMS-HCC risk adjustment model used for MA payments is based only on diagnoses from medical claims.

[51] 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.

individual-specific heterogeneity in the relative coding intensity between plans A and B. For individual i, this generates an "upcoding factor" of $\alpha_B + \epsilon_{iB} - \alpha_A$. This produces an average upcoding factor of $\frac{1}{N}\sum(\alpha_B + \epsilon_{iB} - \alpha_A)) = \alpha_B + \bar{\epsilon} - \alpha_A$.

Letting $\mathbb{1}[B_i(\theta)]$ represent the indicator function for choosing B, market average risk can be expressed as $\bar{r} = \frac{1}{N}\sum(\hat{r}_i + \alpha_A + \mathbb{1}[B_i(\theta)](\alpha_B + \epsilon_{iB} - \alpha_A))$. Assuming that ϵ_i is orthogonal to θ, differentiating $\bar{r}$ with respect to θ produces, $\frac{\partial \bar{r}}{\partial \theta} = \frac{\partial}{\partial \theta}\frac{1}{N}\sum(\hat{r}_i + \alpha_A + \mathbb{1}[B_i(\theta)](\alpha_B + \epsilon_{iB} - \alpha_A)) = \alpha_B + \bar{\epsilon} - \alpha_A$. Since $\bar{\epsilon}$ is constant, $\hat{\alpha}_B$ can be defined as $\alpha_B + \bar{\epsilon}$ so that $\frac{\partial \bar{r}}{\partial \theta} = \hat{\alpha}_B - \alpha_A$, the average upcoding factor.

A.4 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. 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 can be estimated by regressing either the average county risk score within TM or the average county risk score within MA on contemporaneous and lagged penetration.

Table A2 presents the selection results. Coefficients on contemporaneous MA penetration identify pure selection effects. 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 TM average, increasing the average risk score in both pools. In Table A2, 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.[52]

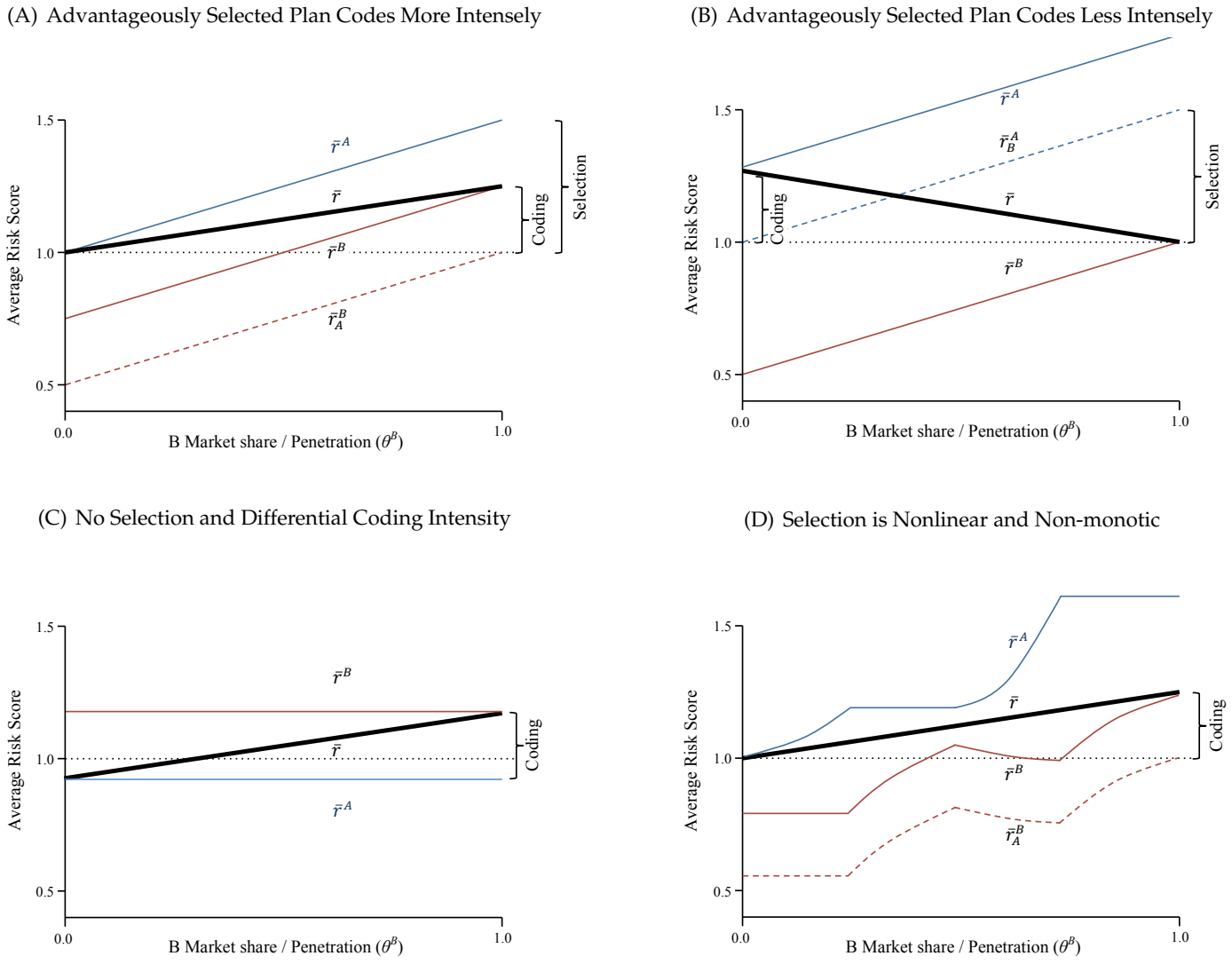
An important component of selection effects may be captured by the lagged penetration coefficient: Research on MA enrollment by Sinaiko, Afendulis and Frank (2013) shows that the majority of new MA enrollees are age 65, implying that most of the shift in MA penetration is likely occurring among the newly Medicare-eligible. In Table A2, this would cause a significant fraction of selection effects to be captured by the lagged coefficient, as new MA enrollees aren't assigned risk scores until their second year.

In short, interpreting selection effects in Table A2 is difficult because of the differential risk-scoring treatment of the newly-eligible 65 year-olds. Coefficients on lagged MA penetration are affected by: (i) selection on risk score trajectory and (ii) selection on 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 lagged coefficients on *overall* county risk, are unaffected by selection and yield a more straightforward identification of pure coding effects.

[52] Regressions in which the dependent variable is the risk score estimate only compensated selection. This is in contrast, for example, to Brown et al. (2014) and Cabral, Geruso and Mahoney (2014) which are primarily interested in uncompensated selection.

Figure A1: Identifying Coding Differences in Selection Markets



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

Table A1: Difference-in-Differences: Coding Across the Age 65 Threshold for Beneficiaries Choosing TM and MA

	Dependent Variable:					
	Risk Score		Count of HCCs		At least 1 HCC	
	(1)	(2)	(3)	(4)	(5)	(6)
Post-65	0.023*	0.023**	0.031	0.031*	0.005	0.005
	(0.011)	0.007	(0.023)	(0.015)	(0.010)	(0.007)
Selected MA	-0.072**		-0.175**		-0.071**	
	(0.013)		(0.027)		(0.014)	
Post-65 X Selected MA	0.044*	0.044**	0.123**	0.123**	0.060**	0.060**
	(0.021)	0.015	(0.043)	(0.030)	(0.020)	(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: Table shows coefficients from difference-in-differences regressions that compare the change in coded diagnoses across the age 65 threshold among consumers who select TM relative to consumers who select MA. The dependent variables are the risk score in columns 1 and 2, the count of HCCs in columns 3 and 4, and an indicator for being diagnosed with a condition yielding at least one HCC in columns 5 and 6. Odd columns include individual fixed effects. Data are as described in Section A.2. Observations are person-years. Standard errors in parentheses are clustered at the person level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Table A2: Selection Results: Effects on within-TM and within-MA Risk Scores

	Dependent Variable:					
	Mean TM Risk Score			Mean MA Risk Score		
	(1)	(2)	(3)	(4)	(5)	(6)
MA penetration t	0.037	0.040	0.040	0.025	-0.024	-0.013
	(0.026)	(0.034)	(0.033)	(0.062)	(0.085)	(0.083)
MA penetration t-1	0.045**	0.030*	0.026*	0.087*	0.116**	0.130**
	(0.013)	(0.012)	(0.012)	(0.040)	(0.040)	(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-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: Table shows coefficients on contemporaneous (t) and lagged ($t-1$) Medicare Advantage (MA) penetration. The dependent variable in columns 1 through 3 is the average risk score among TM enrollees in the market (county). The dependent variable in columns 4 through 6 is the average risk score among MA enrollees in the market (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 A3: 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	0.001	0.001	0.011	0.008	0.007	-0.008	-0.012	-0.011
	(0.015)	(0.019)	(0.019)	(0.015)	(0.018)	(0.018)	(0.024)	(0.033)	(0.033)
MA penetration t-1	0.069**	0.067**	0.064**	0.081**	0.065**	0.064**	0.040**	0.055**	0.050**
	(0.011)	(0.012)	(0.011)	(0.014)	(0.016)	(0.016)	(0.013)	(0.019)	(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-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: Table shows coefficients on contemporaneous (t) and lagged ($t-1$) Medicare Advantage (MA) penetration. The dependent variable is the average risk score in the market (county). Because MA risk scores are calculated using diagnosis data from the prior plan year, coding differences can plausibly affect risk scores only with a lag, and coefficients on contemporaneous penetration serve as a placebo test. Observations are county $\times$ years. 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 Part-D plans. Additional 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: Extended Placebo Tests: Effects of Contemporaneous Penetration and Leads

	Dependent Variable: County-Level Average Risk Score							
available panel years:	2007-2011	2007-2010	2007-2009	2008-2011	2008-2010	2008-2009	2009-2011	2009-2010
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MA penetration t+2 (placebo)			0.044+			0.030		
			(0.023)			(0.036)		
MA penetration t+1 (placebo)		0.017	0.032		-0.005	-0.019		-0.004
		(0.025)	(0.056)		(0.015)	(0.042)		(0.034)
MA penetration t (placebo)	0.001	-0.021	-0.064	0.006	0.003	-0.025	0.011	0.014
	(0.019)	(0.028)	(0.071)	(0.017)	(0.025)	(0.091)	(0.016)	(0.043)
MA penetration t-1	0.064**	0.076**	0.084**	0.041**	0.038+	0.025	0.037	0.052
	(0.011)	(0.018)	(0.022)	(0.015)	(0.022)	(0.038)	(0.032)	(0.090)
MA penetration t-2				0.046*	0.054*	0.048	0.052+	0.100
				(0.022)	(0.024)	(0.041)	(0.031)	(0.061)
MA penetration t-3							0.023	-0.033
							(0.024)	(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-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: Table shows coefficients on future ($t+2$, $t+1$), contemporaneous (t), and lagged ($t-1$, $t-2$, $t-3$) Medicare Advantage (MA) penetration. The dependent variable is the average risk score in the market (county). Because MA risk scores are calculated using diagnosis data from the prior plan year, coding differences can plausibly affect risk scores only with a lag, and coefficients on contemporaneous and future penetration serve as a placebo test. Observations are county $\times$ years. 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. 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: Falsification Test: Effects on Age Distribution, Unconditional on Medicare Status

	Fraction ≥65	Conditional on ≥65				
		65-69	70-74	75-79	80-84	85+
	(1)	(2)	(3)	(4)	(5)	(6)
MA penetration t	0.003	0.013*	-0.006	0.002	-0.008*	-0.001
	(0.004)	(0.006)	(0.006)	(0.005)	(0.003)	(0.003)
MA penetration t-1	-0.004	0.002	0.006	-0.011*	0.003	0.000
	(0.004)	(0.005)	(0.004)	(0.005)	(0.003)	(0.003)
Main Effects						
County FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Additional Controls						
State X Year Trend	X	X	X	X	X	X
County-Year Demographics						
Observations	15,640	15,640	15,640	15,640	15,640	15,640

Note: Table shows coefficients on contemporaneous (t) and lagged ($t-1$) Medicare Advantage (MA) penetration. The dependent variable in column 1 is the fraction of the county population ≥ 65. The dependent variables in columns 2 through 6 are the fraction of the county population in age bins conditional on age ≥ 65. Observations are county $\times$ years. Controls are as described in Table 2. Standard errors in parentheses are clustered at the county level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.