Health Care Spending and Utilization in Public and Private Medicare*

*** Preliminary. Do Not Cite or Circulate without Permission ***

Vilsa Curto, Liran Einav, Amy Finkelstein, Jonathan Levin, and Jay Bhattacharya†

September 2016

Abstract. We compare healthcare spending in public and private Medicare using newly available private claims data from Medicare Advantage (MA) insurers. We find that healthcare spending is 27 percent lower in MA than for individuals in the same county and same risk score enrolled in public, traditional Medicare (TM). Spending differences between MA and TM are similar across sub-populations of enrollees and sub-categories of care. They primarily reflect differences in healthcare utilization. Average prices for an admission to a given hospital for a given diagnosis are virtually identical in MA and TM. We present evidence consistent with MA employing various types of utilization management and encouraging substitution to relatively less expensive modes of care, such as use of primary care instead of specialists, and outpatient rather than inpatient surgery. Geographic variation in healthcare spending is larger in MA than in TM, although geographic variation in hospital prices is lower in MA than in TM.

JEL classification numbers:
Keywords:

*We are grateful to Daniel Salmon and Martina Uccioli for excellent research assistance. We gratefully acknowledge support from the NSF (SES-1527942, Bhattacharya, Einav and Levin), the NIA (R01 AG032449, Einav and Finkelstein; R37 AG036791, Bhattacharya), and the Sloan Foundation (Bhattacharya, Einav, Finkelstein and Levin). The authors acknowledge the assistance of the Health Care Cost Institute (HCCI) and its data contributors, Aetna, Humana, and UnitedHealthcare, in providing the claims data analyzed in this study.

†Curto: Department of Economics, Stanford University, vcurto@stanford.edu; Einav: Department of Economics, Stanford University, and NBER, leinav@stanford.edu; Finkelstein: Department of Economics, MIT, and NBER, afink@mit.edu; Jonathan Levin: Department of Economics, Stanford University and NBER, jdlevin@stanford.edu; Jay Bhattacharya: School of Medicine, Stanford Univesity, jay@stanford.edu.

1 Introduction

A long standing question in economics concerns the appropriate roles of the public sector and private sector in providing services that society has decided are essential. This question comes up in many contexts – education, utilities, transportation, pensions – and especially in healthcare. The United States is unusual among developed countries not only in its share of GDP that is devoted to healthcare, but also in its distinctive mix of public and private health insurance. However, comparisons of healthcare utilization and spending under alternative systems are difficult, since public and private health insurance systems do not typically operate on a similar scale, for the same population, in the same market, or with the same providers.

The specific context of the U.S. Medicare program for individuals aged 65 and over may surmount some of these difficulties. This program presents a rare opportunity for a "side by side" comparison of public and private health insurance systems. The Traditional Medicare (TM) program is publicly provided. Over the last decade, a significant fraction of the US Medicare population has opted out of TM and enrolled in private insurance plans through the Medicare Advantage (MA) program. In MA, private insurers attract Medicare beneficiaries, obtain capitated payments from the government, and offer health insurance in a way that roughly mimics commercial health insurance. Currently about a third of Medicare beneficiaries are enrolled in MA.

Empirical work on comparisons of MA and TM has been hampered by asymmetric data availability: administrative claim-level data from TM is naturally available given the fee-for-service nature of reimbursements, but detailed claim-level information from MA insurers was more difficult to access given the capitation by which they were paid for their enrollees. In this paper, we take advantage of newly available, private claims data from Medicare Advantage plans in 2010 provided by the Health Care Cost Institute (HCCI). We combine it with claim-level data from TM to compare healthcare spending, healthcare utilization, and other outcomes in public and private Medicare. Our data consist of the claims data for three Medicare Advantage insurers (Aetna, Humana, and UnitedHealthcare), which together cover almost 40 percent of MA enrollees.

The key advantage of these data is that they allow researchers to analyze claim-level data in MA – i.e. healthcare spending and utilization based on MA plan payments to healthcare providers – in an analogous manner to the existing and widely used claims data for TM. We estimate average MA healthcare spending per enrollee-month of $647, of which $596 is paid by MA insurers and the rest is owed by enrollees out-of-pocket.

We next compare healthcare spending and utilization between MA enrollees and a comparable group of TM enrollees. Our baseline analysis compares MA enrollees to TM enrollees residing in the same county and with the same risk score (a prediction of their spending in TM based on basic demographics and detailed information on prior health diagnoses). This follows the spirit by which Medicare reimburses MA insurers for their enrollees, which is (broadly speaking) to reimburse based on average insurer spending for a TM enrollee with that risk score in that county.

We estimate that monthly healthcare spending per enrollee is 27 percent lower for MA enrollees than for TM enrollees in the same county and risk score. Applying this estimate to the entire MA population (including insurers not in our sample), this suggests that total healthcare spending in MA in 2010 was about \$27.8 billion lower than it would be for comparable TM enrollees.

The lower spending in MA relative to TM is not concentrated in particular sub-populations or sub-categories of care. We see comparable spending differences across different types of enrollees by age, by gender, and by residence in urban vs. rural counties. We also see comparable proportional spending differences appearing at all quantiles of the spending distribution and in both inpatient and outpatient care. Geographic variation (unadjusted for demographics) in healthcare spending is 25 percent larger in MA than in TM. Areas with higher TM spending show larger proportional spending differences between MA and TM.

Lower spending in MA primarily reflects lower utilization of healthcare; we find essentially no difference in average prices for services in MA and TM. MA enrollees have fewer inpatient admissions, fewer outpatient office visits, fewer skilled nursing facility (SNF) visits, fewer physician visits, and fewer ED visits. They have lower utilization both for services where there are concerns about over-use (diagnostic testing and imaging) and for services where there are concerns about under-use (preventive care). Spending per encounter, however, is similar or slightly higher in MA than in TM. For hospital admissions – where it is feasible to compare unit prices for admissions to the same hospital and with the same diagnosis (DRG) – our findings also point to a lack of quantitatively meaningful pricing differences: on average, prices in MA are 1.5 percent higher than prices for the same hospital and DRG in TM. Geographic variation in hospital pricing is about 20 percent lower in MA than TM.

We find suggestive evidence for some potential mechanisms by which MA insurers may reduce utilization relative to TM. The fact that spending per encounter is slightly higher in MA than TM is consistent with utilization constraints in MA, so that the marginal patient admitted to these types of care is in worse health. Relatedly, we find that MA patients are much less likely to be discharged from the hospital to post-acute care and much more likely to be discharged home than TM patients. We also find evidence of restrictions on the most expensive types of care, possibly including substitution to less expensive alternatives. For example, visits to specialists are much lower (23 percent) in MA than TM, while visits to primary care are only slightly lower (6 percent). Similarly, the probability of inpatient surgery is 8 percent lower in MA than TM while the probability of outpatient surgery is 26 percent higher.

The evidence on potential mechanisms helps alleviate – but does not remove – concerns that differences in average spending between MA and TM reflect differences in expected healthcare spending for individuals who select into MA, rather than a "treatment effect" of MA per se. Our baseline results compare spending in MA to spending in TM for individuals in the same county and with the same risk score. To the extent that county and risk scores are the only variables that could be used in any capitation formula, this difference is a useful summary, which may provide a guide for, say, CMS reimbursement rates. Yet it may partly (or entirely) reflect selection effects whereby MA attracts individuals with lower predicted spending, conditional on risk score and county. We therefore explored how estimates of mean spending differences between MA and TM are affected by more detailed controls for observable differences between TM and MA enrollees, as well as by attempts to adjust for unobserved differences between the two populations using data on mortality differences (conditional on observables) to proxy for differences in expected spending. Without any adjustment, spending is 30 percent lower in MA, compared to our baseline estimate of 27 percent. Our alternative attempts to adjust more finely for observable differences or to try to account for unobservable differences between MA and TM enrollees suggest that spending differences could get as low as 13 percent. While none of our approaches is perfect, we view the totality of the evidence as suggesting that MA reduces healthcare spending relative to what it would be in TM by 10 to 25 percent.

Our findings relate to several literatures. Most broadly, our work is part of the large literature on the relative consequences of public and private ownership. This literature has spanned a range of disparate industries, including education, pensions, electricity, and transportation. In the specific context of healthcare, recent empirical work has emphasized that the private sector may be more efficient than the public sector at setting reimbursement prices for providers (Clemens et al. 2015) and at setting cost-sharing to combat moral hazard (Einav et al. 2016).

Not surprisingly, we are not the first to explore the question of the relative efficiency of public and private health insurance by comparing behavior in MA and TM plans. As noted, a key contribution of the current paper is our access to detailed claims data for a large share of the MA market. Absent such data, prior studies have used a variety of approaches to infer health care utilization and spending differences between MA and TM. These include comparing MA plans' self reports to CMS of enrollee utilization to utilization measures in TM claims data (Landon et al. 2012), analyzing beneficiaries' self-reports of care received in TM and in MA (Ayanian et al. 2013), analyzing hospital discharge data from New York counties experiencing MA exit (Duggan et al. 2015), and inferring cost differences from estimates of demand for MA and a supply-side model of the market (Curto et al. 2014). These papers have tended to find lower healthcare utilization in MA – with estimates ranging from 10 percent to 60 percent.

Our finding that MA prices for hospital admissions are roughly similar to TM prices contrasts with the conventional wisdom that MA prices will be higher than TM prices due to the bargaining power enjoyed by the larger public sector (e.g. Philipson et al., 2010). It also differs from prior findings that TM prices are substantially lower than prices in the private, under 65 market both on the inpatient side (Cooper et al. 2015) and the outpatient side (Clemens and Gottlieb, forthcoming). Likewise, is interesting to compare our price findings to the seminal paper of Cutler et al. (2000) comparing *private* managed care compared to *private* fee-for-service provision of health care. Focusing on heart attack admissions in Massachusetts, they find spending per heart attack admission is 30 to 40 percent lower in private managed care than private fee for service; by contrast, we find spending per heart attack admission is about the same (2 percent higher) for private managed care (MA) compared to public fee for service (TM).

Finally, our paper relates to the large literature on geographic variation in healthcare spending in TM (e.g. Fisher et al. 2003a, 2003b; Skinner 2011, Institute of Medicine, 2013), and a smaller but growing literature comparing geographic variation in TM to geographic variation in commercial insurance i.e., private insurance for the under 65 population (e.g. Philipson et al. 2010, Institute of Medicine, 2013 Cooper et al., 2015). The large geographic variation in TM spending without commensurately better mortality in higher spending areas has widely been interpreted as evidence of inefficiency in TM. The comparison with commercial insurance has suggested that TM and commercial spending both exhibit substantial geographic variation, but that geographic variation in TM primarily reflects variation in utilization while geographic variation in commercial insurance primarily reflects geographic variation in pricing. This has been interpreted as reflecting the lower powered incentives in the public sector relative to the private sector in constraining utilization, and monopsony power in the public sector to constrain prices relative to what the private sector can achieve (e.g. Philipson et al., 2010). Of course, there are other reasons why patterns of healthcare provision for those under 65 may differ from the patterns for the over 65. We consider these same set of facts in the context of Medicare Advantage, which arguably provides a cleaner comparison group to TM for understanding variation under private and public regimes since MA and TM are provided to the same broad population (of individuals over 65 years of age). And in contrast to the comparisons between TM and commercial insurance, we found lower pricing variation in MA than TM.

The rest of the paper proceeds as follows. Section 2 provides some institutional background on our setting. Section 3 describes our data and baseline sample. Section 4 presents summary statistics on healthcare spending in MA and introduces our baseline measurement approach for comparing spending in MA to spending for a "comparable" set of TM enrollees. Section 5 compares healthcare spending in MA and TM, overall and for various categories of people and spending. Section 6 examines differences between MA and TM enrollees in healthcare utilization and in healthcare prices, and examines some potential mechanisms for spending reductions. Section 7 explores alternative approaches to controlling for selection into MA. The last section concludes.

2 Setting and background

The context of our paper is the Medicare Advantage (MA) program, which allows Medicare beneficiaries to opt out of traditional fee-for-service Medicare coverage and enroll in a private insurance plans. The program was established in the early 1980s with two goals: to expand the choices available to beneficiaries and to capture cost savings from managed care. In return for covering enrolled beneficiaries' healthcare expenses, private MA plans receive a risk-adjusted, capitated monthly payment from the Centers for Medicare and Medicaid Services (CMS), which is the federal agency that manages the Medicare program.

There has historically been a tension between the two goals of expanding access to MA and limiting costs (McGuire, Newhouse, and Sinaiko, 2011). Insurers have tended to participate more in periods with higher payments, and to offer plans selectively in areas with higher payment rates. MA plans also enroll relatively healthier beneficiaries, complicating the problem of setting appropriate capitation rates. Several reforms over the last decade have aimed to address these problems by introducing a risk scoring system to adjust plan payments based on enrollee health, and a competitive bidding system that replaced the fixed reimbursement rates used earlier. These changes, combined with an increase in capitation rates set by CMS, have coincided with the expansion of plan offerings and enrollment seen in Figure 1. MA penetration tends to be higher in urban than rural areas; for example, in 2010, MA penetration was 32% in urban counties and 18% in rural counties.

To participate in the MA program, Medicare private plans must contract with a set of healthcare providers and offer at least the same insurance benefits as standard Medicare, which covers inpatient ("Part A") and outpatient ("Part B") healthcare services. They typically provide additional benefits as well, in the form of more generous cost sharing or supplemental coverage of dental, vision, or drug benefits. Medicare beneficiaries observe the MA plan offerings in their county of residence and can choose to enroll in any of the available MA plans during an annual "open enrollment" period every fall. The trade-off they face in choosing between MA and TM is that MA plans typically restrict access to healthcare providers, but provide additional benefits as described above. In 2010, around 73 percent of MA enrollees were in HMO or PPO plans with limited provider networks.

Every year, plans enter into a bidding process, which dictates the benefits and premium associated with each plan that is offered to beneficiaries The precise rules of the way plan bids translate to plan premiums and benefits are somewhat complicated (for a more detailed description, see Curto et al. 2014). We briefly summarize the key features here. Each plan submits a bid b, which should be interpreted as the monthly compensation required by the plan to provide "standard" monthly coverage in the local area in which the plan is offered to an "average" Medicare beneficiary. By "standard" coverage we refer to the standard part A and part B financial coverage offered by TM; MA plans typically offer more comprehensive coverage, but they obtain a separate compensation from CMS for it on top of their bid b; this is known as the "rebate." As will be clearer later, by "average" beneficiary we refer to a beneficiary with an average health risk.

This bid b is then assessed against its local benchmark B, which is set administratively by CMS.[1] In principle the benchmark B is supposed to approximate the counterfactual cost to CMS to cover an "average" beneficiary in that county through TM. In practice, the variation in benchmarks across locations departs somewhat from this principle, presumably reflecting various political economy considerations; on average in our observation period (2010), benchmark rates are higher than corresponding TM costs, and more so in some areas than in others.[2] Overall, the average benchmark across counties (weighted by the number of Medicare beneficiaries) is $836 per enrollee-month, compared to an average TM cost of $793, and this difference is lower in urban counties (benchmark of $866 and average TM costs of $837) than in rural counties ($770 vs. $713). The (weighted)

[1] If $b > B$ the difference is charged as a premium to the consumer. If $b < B$, which is almost always the case empirically, 75 percent of the difference is given to the consumer through the rebate, and 25 percent is retained by CMS.

[2] Indeed, one of the components of the recent Affordable Care Act is to reduce the level of these MA benchmark rates.

correlation between the county benchmark and the (realized) average TM cost in each county is 0.75. However, in our observation period, the vast majority of plan bids are lower than the corresponding benchmarks, making MA plans financially more generous than traditional Medicare, where enrollees can face large out-of-pocket costs.

Capitation payment to insurers for enrolling a given enrollee in a given MA plan depends not only on the plan's bid b but also on the enrollee's risk score r_i, which is proportional to her predicted health care costs in TM over the next year. Adjusting reimbursement for risk score is a key component of CMS's attempt to limit selection into MA by adjusting plan compensation for predictable heterogeneity in healthcare cost across beneficiaries. CMS assigns a risk score to each Medicare beneficiary based on demographic information and detailed claim-level information on chronic disease conditions measured over the previous 12 months. The average beneficiary's risk score is normalized to 1, so that plans obtains compensation of $r_i b$ for covering beneficiary i. Thus, broadly speaking, plan compensation is designed to reimburse an MA insurer for the costs their enrollee would incur – based on their county and risk score r_i – had they remained in TM. This motivates our baseline approach (described below) to comparing utilization and healthcare spending in MA and TM, which is to compare outcomes for MA enrollees to outcomes for TM enrollees who are in the same county with the same risk score.

3 Data and Sample Construction

3.1 Data sources

This project uses data from two main sources: the Health Care Cost Institute (HCCI) and the Center for Medicare and Medicaid Services (CMS). All the data pertain to spending and enrollment in 2010. The Appendix provides considerably more detail.

The HCCI data are the key, novel data in this paper. HCCI is provided with claim-level data from three large MA insurers – UnitedHealthCare, Humana, and Aetna. HCCI pools these data (masking the individual insurers) and makes these data available for research In 2010, these three insurers (whom we refer to as the "HCCI Insurers") covered almost 40 percent of MA enrollees: United was the largest (national market share of 18%), Humana was second (15%), and Aetna fifth (4%) (Kaiser Family Foundation, 2010). The claim-level data reflect claims that these three insurers paid out to healthcare providers. The HCCI data also contain monthly enrollment indicators and some limited demographic information (age bins, gender, and zip code).

The CMS data serve a dual role. One role is in providing parallel claim-level data on those Medicare beneficiaries enrolled in Traditional Medicare (TM). Because TM offers a fee-for service coverage, we essentially observe every healthcare claim made by TM enrollees during 2010. With the various adjustments described later, the TM claims data allow us to form a "benchmark" comparison of healthcare spending and utilization against which we can compare the measures obtained from HCCI.

The CMS data have a second, equally important role. Because CMS pays MA insurers on a risk-adjusted capitated basis, CMS cannot track healthcare spending and utilization of MA enrollees, but it has precise information about the enrollment, demographics, and health risk of MA enrollees. Thus, for the universe of Medicare enrollees we can observe monthly enrollment information in TM (parts A and/or B) or MA, risk score, demographics (zip code, age, and gender), dual enrollment status (in Medicaid and Medicare), and detailed health conditions from the prior year. The CMS data also contain mortality information for MA enrollees as well as TM enrollees (because payments to MA insurers stop the month a beneficiary dies), as well as details on payments to MA insurers by CMS and consumers. The detailed CMS data on MA enrollment and mortality allows us to validate the completeness of our baseline sample in HCCI, and to adjust our comparison to TM spending for the differential demographics, health conditions, and mortality among MA enrollees. The CMS data on payments to MA insurers allow us to construct estimates of total payments to HCCI Insurers, as well as their components.

3.2 Baseline sample

The HCCI data include most, but not all, MA enrollees in the three HCCI insurers. Based on the qualitative information that HCCI obtained from the three participating insurers, it appears that inclusion in the HCCI data was made on a plan-by-plan basis, with "highly capitated plans" left out. That is, insurance plans that pay providers on a capitated basis naturally do not observe specific claims by their enrollees, so they are omitted from the HCCI data. The HCCI data also indicate that it excludes special needs plans (SNPs), which are MA plans for individuals with specific diseases (such as end-stage liver disease, chronic heart failure, or HIV-AIDS) or certain characteristics (such as residence in a nursing home).

Ideally, we would have plan identifiers in the HCCI data, which would allow us to match this information to the plan identifiers in the CMS data, and thus know which MA plans are excluded. However, with the exception of SNPs that are not in the HCCI data and can be identified in the CMS enrollment data, plan and insurer identifiers are omitted from the HCCI data.

Instead, therefore, we rely on the fact that capitated payments to providers are more popular in some parts of the US than in others, and the MA market operates separately across locations. We thus construct our baseline sample by focusing on specific regions, in which HCCI plan coverage appears approximately complete.

We judge the completeness of the HCCI data by comparing enrollment statistics for the HCCI insurers in the HCCI and CMS data. In the CMS data, we know for each MA enrollee whether he or she was enrolled in an MA plan offered by one of the HCCI insurers. This allows us to generate a pseudo HCCI enrollment data set, which covers all enrollees that "should" have been in the HCCI data if no plans were omitted. We then compare enrollee counts in this pseudo HCCI enrollment data and cross validate the actual HCCI data against it. Specifically, we compare enrollee-month counts at the state level across the two data sets, restricting the analysis to individuals who are 65 and over; we do not require individuals to be enrolled for a full year. Overall, this exercise suggests that the HCCI data contain about 80 percent of total MA enrollees for the HCCI insurers, with the "missing" plans (and enrollees) disproportionately concentrated in the Western US.

We define our baseline sample to be the set of 27 states where we have a close to complete sample of HCCI insurers' enrollees, which we define to mean that the count of enrollee-months in HCCI in the state is within 10 percent of the count for the HCCI insurers in pseudo HCCI enrollment data that is derived from CMS data.[3] In practice, in these complete data states, total HCCI enrollment is within one percent of total enrollment in the pseudo HCCI enrollment data, leaving us reasonably sanguine that we have captured the entire set of MA enrollees for these three insurers.

The 27 states in our baseline sample represent about 60 percent of enrollees for the HCCI insurers. The excluded states are disproportionately concentrated in the Western United States. Appendix Table 1 shows the MA share of total Medicare enrollees and the HCCI insurer share of MA enrollees for both our complete data states as well as for the states that are omitted from our baseline sample, and Appendix A provides additional details about the construction of the baseline sample.

Table 1 shows how our baseline sample is constructed, and presents basic demographic statistics from both the CMS and HCCI data. Columns (1) through (3) present CMS data across all plans and states, while column (4) through (6) present CMS data for our baseline sample, which covers the 27 states above and omit enrollees in SNPs and in masked (i.e. very small) zip codes. In each case, we present statistics for all TM enrollees, for all MA enrollees, and then for enrollees in the three "HCCI insurers". Columns (7) and (8) present statistics for the HCCI sample, for the entire sample in column (7) and for our baseline sample in column (8).

We use Table 1 to make several observations. First, comparing columns (1)-(3) to column (4)-(6), the 27 states that constitute the baseline sample do not seem to be very different from the overall sample, making us feel reasonably comfortable that the findings we report throughout the paper are likely to be relevant for states not covered by our baseline sample. Second, comparing column (2) to (3) or column (5) to (6), it appears that the three HCCI insurers attract enrollees that seem reasonably similar to the overall MA enrollees, suggesting that our subsequent findings may apply for the broader MA population. Third, as has been documented elsewhere, MA enrollees are slightly younger and significantly healthier than TM enrollees: their risk scores (which are proportional to their predicted healthcare spending) are about 5-6 percent lower, and their annual mortality rates are almost a third lower. This suggests that a straight comparison of TM and MA healthcare spending would be misleading, motivating the various corrections for selection we describe in the next section. Finally, it is reassuring that, for our baseline sample, the demographics (that we can measure in both data sets) are remarkably similar when measured in the CMS data (column 6) and the HCCI data (column 8); this is what we would expect given our construction of a baseline sample for which the HCCI data should include all relevant MA enrollees.

[3] Although we define our inclusion criteria at the state level, in practice the same inclusion criteria applied at the county level would retain all counties within each of the sttes in our baseline sample.

4 Summary statistics and measurement approach

4.1 Spending and payments in MA

The middle panel of Table 1 reports average total healthcare spending and CMS payments in TM and MA. Throughout, we define healthcare spending as the sum of the insurer spending and any out-of-pocket spending by the beneficiary. Insurer spending is based on observed payment amounts – that is, transacted prices, not list prices. Out-of-pocket spending is the amount owed by the enrollee (due to deductibles and co-insurance).[4] Our measure of total spending is a near-exhaustive measure of all healthcare claims. Specifically it covers several categories of spending: (a) inpatient spending, which is associated with providers identified as hospitals and physicians billing for treatment provided in an inpatient hospital setting; (b) outpatient spending, which also includes home health care and durable medical equipment (e.g., wheelchair rentals); and (c) skilled nursing facility (SNF) spending.[5] Average MA healthcare spending per enrollee-month is \$647 in our base sample (column (8)). Of this, \$596 is paid by the insurers, and \$52 is owed by the enrollee.

The bottom panel of Table 1 reports estimates of payments to MA insurers. Payments to MA insurers for "organic" MA services (i.e. for services covered by TM) are \$776 per enrollee-month in our baseline sample (column (6)).[6] The comparison of insurer MA revenue of \$776 per enrollee-month to the insurer payments to healthcare providers of \$596 (column 8) suggests that net revenues for MA insurers are \$180 per enrollee-month. If this applied to the entire MA population in 2010 (including those outside our sample) it would imply \$21.1 billion in annual (2010) revenue for MA insurers in excess of their spending on healthcare claims. Of course, MA insurers incur additional costs, such as administrative and advertising expenses, which we do not observe in our data.

[4] TM enrollees can purchase supplemental private insurance ("Medigap") to cover some of their out-of-pocket expenses or additional benefits. About half do so. If they do, the Medigap insurer is the primary payer of the "out-of-pocket" amount owed by the beneficiary.

[5] One (small) category of spending which is not in our measure of total spending is hospice care. This is because hospice care is billed directly to CMS even for MA enrollees, so it is observed in CMS data, for both TM and MA and doesn't fully conform to the empirical exercise. In practice, we show below that the exclusion of hospice spending does not substantively affect the comparison of total spending.

[6] We define payments to MA insurers to be the sum of CMS spending on MA (\$787) and additional consumer premiums for MA (\$6) minus the portion of the consumer rebate that is passed on to consumers for additional services, not covered by Medicare Part A and Part B services (\$17). As discussed in Section 2, MA insurers typically offer more comprehensive coverage than TM, but they obtain a separate compensation from CMS for it, on top of their bid. On average in our baseline sample, the consumer rebate was \$53 per enrollee-month, and \$36 of it is for more generous coverage of the healthcare services we study in the paper, while the remaining \$17 of the rebate is for additional consumer benefits that are not cpatureed by the analogous TM spending (such as premium discounts, or dental and vision coverage).

4.2 Spending in MA and TM: Raw Comparisons

Table 1 reveals dramatic differences in total healthcare spending between the TM and MA populations. In our baseline sample, the average TM enrollee spends $924 per month (column (4)), while the average MA enrollee spends 30% less, $647 (column (8)).

Figure 2 shows spending differences in MA and TM separately for each of the 27 states in our baseline sample. Spending is lower in MA in all states, but the differences ranges from about 4 percent lower MA spending in Alaska to almost 50 percent lower MA spending in Florida.

Geographic variation in spending within TM has attracted a great deal of attention. The "Dartmouth Atlas" findings of large differences across areas in TM spending and utilization without corresponding differences in mortality is widely viewed as indicative of the inefficiencies of the public Medicare system.[7] Our analysis suggests that if anything, geographic variation in spending is higher in MA than TM. The coefficient of variation across states (weighting each state by its total Medicare enrollment) is 0.133 in MA, about 25 percent higher than the 0.107 coefficient of variation we estimate in TM.[8] In Appendix Figure 1 we show that MA also exhibits the positive correlation across states between spending and mortality that has been widely documented in TM.

4.3 Measurement

The 30 percent lower baseline spending in MA relative to TM may partly or entirely reflect differences in the sets of beneficiaries who enroll in TM and MA, rather than the impact of MA on healthcare spending. Indeed, we have already seen in Table 1 that MA enrollees tend to be healthier than TM enrollees, suggesting that some of the spending differences we observed likely reflect these health differences. This motivates our baseline empirical strategy in which we follow a reweighting specification which makes the TM population look like the MA population in terms of county and risk score. The risk score is a summary statistic based on an extremely rich set of demographic and health measures. These health measures reflect both patient health and propensity to receive healthcare - since diagnoses are only recorded if care is received (Song et al. 2010; Finkelstein et al. 2016) - both of which may differ between TM and MA enrollees.

Specifically, consider a Medicare enrollee in county z_i with (continuous) risk score r_i, and an outcome y_i^{TM} in TM. We map r_i to a discrete risk score bin r_i', so that all Medicare beneficiaries are partitioned into a set of discrete groups, defined by their county and risk score bin $g_i = (z_i, r_i')$. Using the sample of beneficiaries who are enrolled with the HCCI insurers, we assign each group g a weight $w_g = N_g/N$, where N_g is the number of enrollees that belong to group g (in the sample of column (6) in Table 1) and $N = \sum_g N_g$. Each unweighted TM outcome

$$\bar{y}^{TM}_{unweighted} = \frac{1}{N_{TM}} \sum\nolimits_{i \in TM} y_i^{TM} \tag{1}$$

[7] Skinner (2011) provides an overview of this large literature.

[8] Our analysis is at the state level rather than the Hospital Referral Region (HRR) level that is more typical in this literature. This is because many HRRs cross state boundaries and our baseline data are limited to a subset of states.

is then replaced with a reweighted TM outcome

$$\overline{y}^{TM}_{re-weighted} = \frac{1}{\sum_{i \in TM} w_{g_i}} \sum_{i \in TM} w_{g_i} y_i, \text{[9]} \tag{2}$$

which we compare to the corresponding MA outcome

$$\overline{y}^{MA} = \frac{1}{N_{MA}} \sum_{i \in MA} y_i^{MA}. \tag{3}$$

In addition to the transparency and simplicity of this re-weighting approach, it has the added attraction that it captures the spirit by which MA insurers are being paid by CMS. As described in Section 2, CMS capitation payments to MA insurers are based on a county-specific amount, which is then multiplied by the enrollee's risk score r_i. Our baseline approach, which reweights on precisely these two dimensions – county and risk score – can therefore be viewed as correcting for selection concerns associated with the only two dimensions that current CMS payment policies condition on.

Naturally, however, there may still be selection into MA on characteristics which, conditional on risk score and county, are correlated with expected health care spending. In Section 7 we return to this fundamental issue, and report several alternative strategies for adjusting for selection into MA using both a richer set of observables (implemented via propensity-score matching) and an attempt to account for selection on unobservables using observed mortality differences between MA and TM enrollees. As we show there, while various approaches to selection move some of the numbers around, the qualitative and the ballpark quantitative conclusions do not change dramatically in most specifications relative to our baseline approach. This makes us comfortable "riding" this relatively simple baseline approach for much of the paper.

Table 2 shows how the TM spending benchmark is affected by different ways of reweighting the TM enrollees to "look like" the MA enrollees in terms of county composition and risk score. Column (1) reproduces the raw, unweighted numbers already shown in Table 1, column (4). Column (2) reweights the TM data to match the distribution of MA enrollees across counties. Average TM spending per enrollee month increases from \$924 to \$950, reflecting the fact that MA enrollees are disproportionately in more expensive counties; this is primarily driven by the well documented higher MA penetration in urban areas, in which healthcare delivery tend to be more expensive. Columns (3) and (4) add risk scores to the reweighting of the TM population, so that it matches, county by county, the risk score distribution of MA enrollees. In column (3) we match on risk score bins that are quite coarse, of width 0.5; 58% of MA enrollees are in the three largest bins (0.5-1, 1-1.5, and 1.5-2). In column (4) we use more granular risk score bins (of width 0.1). It is evident from column (3) (and not surprising given Table 1) that reweighting on risk scores is

[9] A slight complication of this procedure arises when an MA enrollee belongs to a group for which there are no TM enrollees, which may happen in small counties and high risk scores, which are less common. This applies to only 0.07 percent of enrollee-months. In such a case, we amend this procedure with an extra step, where we re-classify to such "empty" TM groups the TM enrollee in the same county whose risk score is the closest to the corresponding unmatched MA enrollee.

important, reducing the average monthly spending by 6.5% relative to reweighting on county only in column (2). However, it is quite remarkable that the much more granular matching on the risk score distribution makes little difference, with columns (3) and (4) showing essentially identical results. Going forward, we will use the re-weighting strategy in column (4) – using county and risk bins of width 0.1 – as our baseline when reporting mean spending or quantity differences between MA and TM. We will show both unweighted and reweighted statistics throughout the paper, but will concentrate our discussion on the reweighted statistics.

5 Differences in Spending in MA and TM

Overall differences Table 2 shows average spending differences across all our baseline sample enrollees in MA (column (5)) and comparison samples in TM. We focus our discussion on our baseline re-weighting strategy in TM (column (4)). This comparison indicates that healthcare spending by MA enrollees is $237 per month (27%) lower than a comparable (on county and risk score) sample of TM enrollees. By comparison, the unweighted data indicates that MA spending is $277 (30%) lower than TM enrollees.

Recall that for our baseline sample, CMS pays MA insurers $787 per coverage month for each MA enrollee (see Table 1, column 6). This is an additional $38 per enrollee month above the $749 that CMS pays for a comparable TM enrollee (see Table 2, column (4)).

Stated differently, in the spirit of CMS' capitation payment formula, if total healthcare spending of MA enrollees under TM was the same as TM enrollees with the same risk scores in the same counties, they would cost $884 per coverage month, while in MA their total healthcare spending is only $647. Applying this estimate to entire MA population in 2010 (column (2) of Table 1, which includes those outside of our baseline sample) this translates to $103.8 billion in annual (2010) healthcare spending in TM relative to $76 billion in healthcare spending in MA, a difference of $27.8 billion in annual healthcare spending.

Differences by consumer type Panel A of Table 3 reports the spending differences for different types of enrollees. Each row represents a different subsample of enrollees. Across the board, overall spending in MA is always significantly lower than the (re-weighted) TM analog; the average difference reported in Table 2 is not driven by specific sub-population. Yet, we see some heterogeneous effect across types of enrollees. The percentage difference in spending is somewhat higher for females than for males. The difference is higher in both absolute and relative terms for elderly beneficiaries. The youngest Medicare beneficiaries (aged 65-74) are associated with a lower MA spending of $134 per month (20%) while the most senior (85 years old and over) are associated with a difference of $409 (31%) per month. The 75-84 group is in between. Looking at beneficiary location, the spending difference is much greater for urban counties, which is where the vast majority (81%) of MA beneficiaries enroll. In urban counties, MA spending is 29% lower than TM spending, while in rural counties it is only 15% lower. Put differently, average spending per month in MA is almost the same for rural and urban counties, but TM spending is much higher in urban counties, thus generating the differential difference. This sharp difference between urban and rural counties is also reflected in the MA revenues (i.e. in plan payments for "organic" MA services from Table 1, panel C), which we estimate to be $208 higher than claims cost in urban counties and only $76 higher in rural ones.

Panel A also reveals an interesting aspect of the role that the reweighting adjustment. A comparison of columns 2 and 3 reveals that reweighting does not reduce the monthly TM spending estimates uniformly across different sub-populations. Using beneficiary age to illustrate, we note that the re-weighting adjustment almost makes no difference for the most senior (85 and older) – essentially suggesting that there is little systematic selection on county and risk scores for this subgroup (or that if there is, it cancels out) – but a larger difference for younger beneficiaries.

Panel B of Table 3 reports results for the realized spending distribution, comparing different quantiles of the MA and TM spending distributions. This allows us to assess whether the spending difference is driven by, e.g., the highest spenders. Again, we see the overall lower MA spending across all parts of the distribution. We see a larger percentage difference at the lowest end, a fairly stable (and sizeable, of 30-35%) difference throughout much of the distribution, and then a somewhat lower percentage difference at the very top one or two percentiles.

Figure 3 shows that states with higher TM spending have greater MA "savings" as measured by the percentage difference between MA spending and adjusted (i.e. re-weighted in risk score and county) TM spending. This is consistent with the "conventional wisdom" that higher spending TM areas are less efficient or productive (Skinner 2011).

Differences by spending type Table 4 looks at spending differences across different categories of care. It shows spending broken down into three mutually exclusive and exhaustive categories: inpatient, outpatient, and SNF. MA spending is lower in all three categories. It is 20% lower for inpatient, 27% lower for outpatient. There is a much larger difference in SNF spending, where MA spending is 50 percent lower than in TM. However, SNF spending accounts for only a small share (11%) of overall spending, so this large percentage difference does not contribute much to the overall difference in spending. We return to the SNF results when we discuss potential mechanisms for reducing healthcare use in Section 6.3 below.

Table 5 compares geographic variation in each component of spending in MA and in (unadjusted for demographics) TM spending. All three components of spending exhibit greater geographic variation in MA than in TM. But the differences are particularly pronounced for SNF spending, where the coefficient of variation in MA is 0.34, compared to 0.18 in TM. The Institute of Medicine (2013) recently called attention to the fact that variation in post-acute spending is a major driver of geographic variation in TM spending. This appears to be true in MA as well, where the geographic variation in SNF spending is even larger (relative to other types of spending) than in TM.

The bottom row of Table 4 reports hospice spending in MA and TM. As noted earlier, hospice is covered by TM for both MA and TM enrollees. It is therefore not in our HCCI data on MA spending and we do not include it in our baseline "total spending" measure. It is however captured

– for both MA and TM enrollees – in the CMS data, which we use to construct spending for both TM enrollees and enrollees in the three MA insurers in the HCCI data. Because MA insurers do not bear the cost of hospice expenditures, they might have an incentive to steer patients to hospice, so that some of the lower MA spending in inpatient, outpatient, and SNF could be offset by higher spending in hospice. The bottom row of Table 4 suggests, however, that this is not the case. Hospice spending is too low to have any potential significant offset effect; moreover, it is also lower (rather than higher) for MA enrollees than for TM enrollees.

6 Differences in utilization, not in prices

In the last section we found a substantial difference, of 27%, between the overall spending in MA and the overall spending of a comparable set of TM enrollees. A natural question to ask is whether this large spending difference is driven by lower healthcare utilization in MA or by the ability of MA insurers (at least the large ones, from which we have data) to negotiate lower prices, or both. This is the focus of this section. One challenge throughout this section is to conceptually separate prices from quantity or quality of care, and this challenge dictates some of the exercises we report. To preview our results, we find that quantity differences appear responsible for the entire difference; various measures of "prices" are all quite similar in MA and TM.

6.1 Differences in encounters and spending per encounter

Table 6 measures various components of health care utilization (Panel A) and for many of these measures a parallel comparison of spending per encounter (Panel B). As already mentioned, encounters could be different – providers may be of different qualities, patients may have different needs and diseases, etc. – so comparing "spending per encounter" is not the cleanest measure of price. Yet, the sharp differences in the comparison is quite indicative, and the results remain qualitatively similar when we focus below on more granular units of healthcare for which spending can be more cleanly considered as "price."

Specifically, in Panel A of Table 6 we present several rough units of utilization. As before, all measures are per enrollee-month and we focus our discussion on the comparison between MA utilization (column (3)) and the reweighted TM utilization (column (2)). Across all categories, utilization in MA is substantially lower. Inpatient admission rates are 18% lower. Conditional on an inpatient admission, length of stay is also slightly (6%) lower in MA, so that overall inpatient days are lower in MA by 23%. This is quite similar to the difference in inpatient spending (of 21%) presented earlier in Table 4. Likewise, total SNF days are more than 50 percent lower in MA, which again is quite close to the 50% difference in SNF spending shown earlier in Table 4.

We also measure outpatient utilization. Rates of outpatient emergency room (ER) visits – that is, ER visits that do not result in an inpatient admission – are 16.5% lower in MA, which is quite similar to the 17.8% difference in inpatient admission rates. We see a similar difference (of 18%) in the number of (outpatient) physician visits per month; this difference is approximately equally driven by the extensive and intensive margin: a 11% lower rate of MA enrollees who see physician at least once during the month and a 8% lower average number of physician visits by MA enrollees who visit the physician at least once.

Given the close similarity between the percentage difference in the above utilization measures and the corresponding differences in spending, it is not surprising that "spending per encounter" – shown in Panel B of Table 6 – is quite similar between MA and TM. Inpatient spending per admission and inpatient spending per day are essentially the same in MA and TM, and SNF spending per SNF day is 2 percent higher in MA. Interestingly, spending per outpatient ER visit is 10 percent higher in MA; this may reflect utilization management for MA patients that discourages relatively less severe cases. We also note that the reweighting approach makes little difference in Panel B; the spending per encounter are quite similar already in the raw comparison of means.

We conclude this section by briefly considering a case study: admissions for AMIs. The treatment of AMIs has received considerable attention in the health economics literature in general, and in the study of managed care in particular. Cutler et al. (2000) compared treatment for heart attack patients in a *private* managed care (HMO plans) and *private* FFS plans in Massachusetts, and found that spending per episode was about 30 to 40 percent lower for managed care, but that treatments were similar, concluding that the differences in spending per episode reflect differences in prices for similar services. The last two rows of Panel B show that spending per AMI admission is about 2 percent higher in MA (and spending per day is about 2 percent lower).[10]

6.2 (Lack of) Mean price differences for hospital admissions for specific diagnoses

Table 6 provides a clear indication that much of the difference in spending is driven by lower utilization in MA. However, to have a cleaner assessment of price differences it is important to compare spending for "identical" units of care. That is, MA and TM enrollees may see different physicians or visit different hospitals, so that comparisons of spending per visit or admission may not reflect differences in the price for the same bundle of services. Similarly, perhaps MA enrollees are disproportionately admitted for more severe diagnoses (DRGs) that require more care, so comparing per-admission spending across the two populations may confound price differences and DRG composition (which was one motivation for looking at the AMI "case study").

We therefore hone in on "pricing" – or unit payment rates – by focusing on the payments in MA and TM for admission to the *same* hospital with the *same* DRG. In this analysis, differences in prices will not reflect differences in providers or in the reason for admission. We focus on the approximately 4,000 hospitals in our baseline sample that are paid (for TM) under Medicare's

[10] This analysis of spending per admission for a given condition is similar in spirit to Baker et al.'s (2016) analysis of spending per admission for a common basket of DRGs and geographic areas. They also use HCCI data (from 2009 and 2012) and, focusing on large DRGs and large metropolitan areas, conclude that MA spending per admission is 8 percent lower than TM spending for the same "basket" of DRGs and areas.

prospective payment system (PPS). These represent about 95 percent of all inpatient admission in MA and cover essentially all standard (non specialty) hospitals. In the context of TM, such hospitals are paid by CMS based on a pre-set formula that is a product of a hospital-specific rate and a DRG-specific rate; it is our understanding (although no such contractual data is available to verify it) that these hospitals are predominantly paid by MA insurers in a similar way. In TM, and presumably in MA as well, some accommodation for exceptions is allowed, resulting in payments that may deviate from the DRG-hospital formula rates.

We compute an average price per admission for each DRG. Appendix C provides more details of how we compute these prices. In computing DRG-specific average prices, we weight the admissions in each DRG by the state's share of MA admissions in all DRGs, so that any differences in average prices across DRGs reflect price differences for a common "state basket," and are not contaminated by differences in the geographic distribution of admissions by DRG across states. The national average price is computed by weighting each DRG by its (national) share of MA admissions.

We compute a parallel set of prices in MA and TM. For both, our starting unit of analysis is an admission in MA, which is characterized by a hospital and a DRG. The MA price is simply the observed (transacted) payments for the admission in the MA claims data. Construction of the TM price proceeds in two steps.[11] First, for each MA admission, we calculate the formula price in TM, applying the PPS reimbursement formula, which is simply the product of a hospital-specific "base" payment rate times a diagnosis-specific (DRG) weight; both are publicly available from CMS. As noted, there are various exceptions that result in payments that may deviate from the formula-based TM payment we constructed. Because we are comparing TM pricing to actual (transacted) payments in MA, in our second step we adjust our TM formula prices to reflect average differences between formula and actual prices. We do so by using the CMS data to estimate a proportional DRG-specific "adjustment factor", which is essentially the ratio of observed TM payments to the TM formula prices we calculated. We multiply the average TM formula price in that DRG by our DRG-specific adjustment factor, to arrive at our final, average TM price in that DRG.

Figure 4 shows the average prices in TM and MA overall, and for the top 20 DRGs (by their share of MA admissions). Appendix Table 2 shows the underlying numbers. The national average admission price in MA is $10,084; in TM, it is $9,927. In other words, consistent with the suggestive evidence from Table 6, our pricing analysis finds very little difference between MA and TM average prices. The price for an average MA admission is only 1.5 percent higher in MA relative to TM. The largest difference among the top 20 DRGs is for a major joint replacement (DRG #470), for which average MA price is about 4% higher than TM. For 12 of the top 20, the average price in

[11] In principle, we could follow the exact same approach as for MA prices, since in the CMS data we observe TM payments for each admission, along with its hospital and DRG. In practice, however, we are constrained from implementing this directly both because hospital identifiers are encrypted in the MA data, and because our DUAs prohibit our exporting data below a minimum cell size. Fortunately, the TM hospital-specific base payment rates (which determine the TM formula payments) are available in our MA data; we are extremely grateful to Zack Cooper for providing us with this mapping.

MA is within one percent of that in TM, and for the other 7 it is within 2.5%.

The close similarity of inpatient admission prices between MA and TM is interesting given that it is frequently conjectured that because the public sector has greater bargaining power, public fee-for-service may achieve lower prices than private insurance (e.g. Philipson et al. 2010). Consistent with this conjecture, prior empirical work has shown that for the same service, TM tends to reimburse at substantially lower prices than commercial (under 65) private insurance both in the outpatient setting (Clemens and Gottlieb, forthcoming) and the inpatient setting (Cooper et al. 2015). In contrast, we do not find that TM prices are substantially lower than privately provided Medicare prices.[12]

Geographic variation in hospital prices In addition to means, we also compared geographic variation in inpatient prices for MA and TM. Figure 5 shows the results; Appendix Table 3 shows the underlying numbers. We construct average state prices in MA and TM following a parallel process to what we did for measuring DRG prices; now, we weight the admissions in each state using the DRG's national share of MA admissions, so that comparisons of state-level average prices are not contaminated by differences in the mix of DRGs across states.[13] Weighting by Medicare enrollment in the state, pricing variation across states in MA is about 20 percent lower than in TM. Specifically, the coefficient of variation across states is 0.063 in MA, compared to 0.077 in TM. By contrast, recent work has shown evidence of substantially higher geographic pricing variation in commercial (less than 65) private plans compared to TM (Philipson et al. 2010, Institute of Medicine 2013, Cooper et al., 2015).[14]

6.3 Specific aspects of utilization differences and potential channels for saving

Our results thus far strongly point to differences in utilization metrics, rather than payment rates, that are driving the overall differences in spending between TM and MA. Here we focus on several

[12] Of course, our MA sample is limited to three large insurers, and their bargaining power may not be representative of smaller MA insurers; on the other hand, Cooper et al. (2015)'s analysis of commercial pricing was also limited to the same three large insurers, and there average inpatient prices were almost twice as high than in TM.

[13] Likewise the adjustment factor we apply to the TM formula average price in the state is now state-specific rather than DRG-specific.

[14] Like us, this analysis focuses on pricing variation in hospitals. The recent Cooper et al. (2015) comparison of pricing variation in TM compared to commerical (i.e. private, under 65) plans also uses data from HCCI, specifically 2007-2011 data for commercial insurance. We confirm that using data only from 2010 and from the subset of 27 states in our baseline analysis and our approach to measuring pricing, we replicate their finding of substantially greater variation in pricing in commercial insurance relative to TM. Specifically, again using the MA share of admissions in each DRG to construct average prices for each state, and estimating the coefficient of variation across states weighting each state by the Medicare enrollment in that state (as in Figure 5), we estimate that pricing variation is over 50 percent larger in commercial insurance (coefficient of variation = 0.13) than in TM (coefficeint of variation = 0.08).

specific, narrow metrics of quantity to try to gain a better understanding of the nature of and potential channels for these quantity differences.

Over-used and under-used care In Table 7 we explore differences in potential low-value and high-value care. Panel A examines utilization of diagnostic testing and imaging services,where excessive use may be a concern (e.g. Brot-Goldberg et al. 2015, U.S. Government Accountability Office, 2008). Panel B examines utilization of various measures of preventive care, an area where under-use may be a concern.

The results show no evidence that one type of care exhibits greater differences between MA and TM than another. Diagnostic tests and imaging procedures are, respectively, 27% and 20% lower in MA, which is similar to, and not higher than, the percentage difference in total spending. Preventive care also exhibits no obvious pattern relative to overall care.[15] Rates of most preventive care are lower in MA, although there is variation across the measures. Flu shot rates for MA beneficiaries are much lower (by 38%). Most other preventive tests are also lower in MA but the differences are smaller. Interestingly, the one area where screening tests are done more frequently in MA than TM is screening tests that are female-specific (mammograms and pap smears), for which the rates are a little higher in MA.

Potential channels for reducing utilization Potential mechanisms by which MA plans may reduce care utilization include: limited provider networks through which beneficiaries receive care, coordination of care programs to more efficiently deliver appropriate services and avoid excessive utilization, and financial incentives to physicians to influence the quality and quantity of services delivered (e.g. Landon et al. 2012). By contrast, in TM there are virtually no restrictions on physician clinical decisions or patient choices of care; TM provides fee-for-service reimbursement to providers and no explicit utilization oversight or controls

We have already seen evidence of one "signature" of these mechanisms: all these mechanisms should constrain entry into care, particularly expensive care, so that the average person using that care in MA is in worse health, and higher cost than the average person using that care TM. In other words, MA enrollees should have fewer encounters, but have greater spending (or utilization) per encounter. Consistent with this, we found that spending per inpatient day, spending per SNF day, and especially spending per outpatient ED visit were all higher in MA than in TM (Table 6, panel B).

In Table 8 we provide additional evidence consistent with restrictions on utilization. In Panel A we explore differences between TM and MA in the distribution of discharge destinations of hospitalized patients. Destinations are roughly ordered in how expensive they are (from cheaper to more expensive). The number of enrollee-months sent to different destinations are uniformly lower

[15] We show rates of preventive care by enrollee-month to be consistent with the analysis in the rest of the paper. Naturally, recommended care is not at a monthly level but typically at an annual (or bi-annual) level. The analysis looks similar if instead we examine these measures on an annual basis (not shown).

in MA, reflecting the lower total number of inpatient admissions in MA (see Table 6). However, the results indicate that patients covered by MA are disproportionately discharged to less expensive destinations. Discharges of MA enrollees directly to home is only 12% lower than TM, discharges to a home health organization is 21% lower, and to SNF it is 37% lower; together, these three destinations make up about 85 percent of discharges in either TM or MA. Discharges to other post acute institutions (such as long-term care hospitals, cancer centers, or psychiatric hospitals) are less common, but significantly more expensive; they are 50% lower in MA than TM.

Finally, we consider substitution to less expensive alternatives. In addition to limiting use of care, MA may also constrain the type of service, encouraging use of less expensive substitutes. Panel B of Table 8 points to some patterns that are suggestive of such channels. First, we analyze the frequency of surgeries. We find the surgery rate to be in fact higher, not lower, in MA by a fair amount (18%). However, inpatient surgeries are lower (by 8%) and outpatient surgeries are much higher (by 26%), which is suggestive of MA insurers using outpatient surgeries to substitute away from inpatient surgeries or perhaps also (given the fact that overall number of surgeries is higher) from other type of expensive, non-surgical admissions. Second, we examine two types of physician visits: primary care and specialist visits. We already saw in Table 6 that MA enrollees are associated with 18% fewer physician visits. While, consequently, both types of physician visits are lower in MA, the percentage difference in the number of specialist visits is much greater. Primary care visits is only 6% lower, while visits to specialists are 23% lower.

7 Alternative approaches to correct for selection

The results thus far suggest that spending in MA is more than 25% lower than in TM, even after adjusting for county and a detailed measure of predicted health spending. To the extent that county and risk scores are the only variables that could be used in any capitation formula, this difference is a useful summary, which may provide a guide for, say, CMS reimbursement rates.

Nonetheless, we would like to know the extent to which this lower spending reflects a treatment effect of MA as opposed to selection into MA by individuals who – conditional on risk score and county – have lower predicted spending due to unmeasured differences in health or preferences for healthcare. The relative importance of selection or treatment is particularly important in the context of assessing the cost implications of any expansion of the MA program to cover those currently enrolled in TM.

We take several steps in this section to try to make progress on this question. We begin by showing that our baseline comparison of mean MA and TM spending is not sensitive to alternative, and arguably richer ways of controlling for observables. We then consider the possibility of selection on unobservables, by using differential mortality rates (conditional on observables) in MA and TM to proxy for unobservable spending differences. This approach yields varying results, with the most conservative suggesting that MA spending is only 13 percent lower than what it would be under TM.

The rest of this section discusses the implementation and results in detail. We emphasize at the outset that we consider these alternative approaches useful but clearly not a panacea for concerns about selection. Our earlier evidence pointing to potential channels by which MA may reduce spending – such as via substitution from inpatient to outpatient surgery or from specialist care to primary care – complements our empirical exercise here in suggesting the existence of an MA treatment effect. One would need a subtler selection story than simply selection into MA based on predicted spending, to explain these patterns. The same is true for many of our other results, such as the comparison of geographic variation. Overall, we view the results as pointing to a large "treatment effect" of MA on spending, in the range of 10 to 25 percent.

7.1 (Standard) framework

A standard potential outcome framework is useful to organize our exercise. Let $W_i = 1$ if beneficiary i is enrolled in a plan offered by one of the three HCCI insurers in MA, and $W_i = 0$ if i is in TM. Let y_i^{TM} be the individual outcome of interest (e.g., healthcare spending per month, which is focus on in this section) if she were in TM, and y_i^{MA} be the individual outcome of interest if she were in MA. We observe $y_i = y_i^{TM}$ when $W_i = 0$, and we observe $y_i = y_i^{MA}$ when $W_i = 1$ The individual treatment effect is $\Delta y_i = y_i^{MA} - y_i^{TM}$.

We observe (e.g., in Table 1, Panel B)

$$D = E\left[y_i^{MA}|W_i = 1\right] - E\left[y_i^{TM}|W_i = 0\right] = T + S, \tag{4}$$

where T is the average treatment effect for the MA population[16]

$$T = E\left[y_i^{MA} - y_i^{TM}|W_i = 1\right] \tag{5}$$

and S represents the selection effect, given by

$$S \equiv E\left[y_i^{TM}|W_i = 1\right] - E\left[y_i^{TM}|W_i = 0\right]. \tag{6}$$

A key advantage – in the context of our data – of the above representation of the selection effect is that it is only a function of y_i^{TM}; this is attractive because the set of observables is significantly richer and more granular in the CMS data than in the HCCI data, and the above representation allows us to analyze the selection effect using CMS data alone, holding the average outcome of interest fixed in the HCCI data.

7.2 Correcting for selection on observables

Our baseline re-weighting approach (see equation (2)) can be viewed within this framework, as assuming that conditional on county and risk score W_i is as good as random assignment. The risk

[16] While there is undoubtedly heterogeneity across MA insurers, our data do not permit analyses of separate effects by insurer.

score itself is generated from an underlying much richer set of observables, including very detailed health measures as well age, gender, and dual eligibility in Medicaid. These observables are used with a particular functional form to produce the risk score. A more flexible approach therefore would be to condition on the individual components of the risk score.

If we want to condition on a richer set of observables, it gets more difficult to apply our baseline re-weighting strategy as the data becomes sparse and it becomes common to observe MA beneficiaries with a vector of characteristics for which there is no match in the TM sample. We therefore instead follow a standard approach of constructing propensity scores for enrollment in MA as a function of a rich set of observables, and then apply the reweighting strategy to the propensity score rather than to the entire vector of observables.

Specifically, given a vector of observables x_i we estimate a logit model of W_i on x_i. That is, we assume that $p_i = \Pr(W_i = 1) = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}$ and estimate β by maximum likelihood. We estimate the logit model separately for each county, to allow the relationship between enrollment in MA and observables to differ across counties. We then use our estimate of β to generate the propensity score for individual i, denoted by $\widehat{p_i}$. Appendix Figure 2 shows the distribution of propensity scores under MA and TM for our baseline $x_i's$ (county and risk score).

We then apply the same reweighting procedure used earlier, in two ways. First, we simply reweight the TM sample to match the propensity score distribution of the MA sample by binning the propensity score to bins of size 0.01. Second, because we think that location heterogeneity may be particularly important for healthcare spending, we also reweight on county and propensity score, so that reweighting groups are defined as a combination of county and a 0.01 bin of propensity score. Table 9 reports the results from both procedures, in columns (3) and (4), respectively, but it turns out that the additional conditioning on county in the reweighting procedure makes little difference, presumably because all our specification already include county-specific estimates in the construction of the propensity score.

Rows 1 and 2 of Table 9 report the TM average spending when we apply no weights (row 1) and when we reweight on risk score (row 2), as in the baseline specification, regenerating the results from Table 2.[17] Rows 3 through 6 report specifications based on the propensity score reweighting. Row 3 shows that (not surprisingly) using county and risk score as in our baseline approach, but through a propensity score has essentially no effect on the results. Rows 4 through 6 then use the individual components of the risk score separately as observables x_i in the propensity score; specifically we now include increasingly flexible controls for age, gender, dual eligibility, and 70 hierarchal condition categories (HCC) indicator variables. These alternative specifications have very little effect on the results. Our lowest estimate of TM spending is \$873 per enrollee-month (relative to the \$884 in our baseline specification), which still implies a 26% lower spending in MA (instead of 27% in our baseline specification).

[17] To be parallel with the subsequent propensity analysis, column (4) of row 2 shows our baseline estimate which re-weights on county and risk score bin, while column (3) shows the impact of re-weighting only on risk score bin.

7.3 Using mortality to correct for selection on unobservables

Although our baseline results are not sensitive to controlling for an increasingly rich set of observables, one is naturally still concerned about the possibility that selection into MA is correlated with unobservables that are correlated with healthcare spending. We try to address selection on unobservables by leveraging the fact that, fortunately, we can observe mortality outcomes for individuals in both TM and MA. As we saw in Table 1, mortality is lower for MA enrollees than TM enrollees; it is also lower conditional on county and risk score (not shown).

We therefore use mortality differences across enrollees in MA and TM to try to proxy for unobservable differences in expected health care spending. We make the strong assumption that mortality outcomes are unaffected by enrollment in MA. Under this assumption, we can use mortality as an additional observable and control for it.

We use differences in mortality rates to create an MA-specific mapping and a TM-specific mapping from risk scores to a health index, captured by mortality. This approach therefore captures differences in expected spending, conditional on risk score, arising from unobservable health differences captured by mortality; this includes not only selection on unobserved health, but also potential upcoding (manipulating of patient diagnoses and hence risk scores) in MA (Geruso and Layton 2015).

Specifically, we predict mortality rate as a nonparametric function of risk score, separately for MA and TM enrollees. We can then reweight on predicted mortality (row 7 of Table 9) or on a propensity score that uses predicted mortality as the observable (row 8 of Table 9); here we once again estimate the propensity score county by county. These specifications lead to the largest estimate of selection. According to these estimates, accounting for selection, total spending in MA is only 13% lower than in TM ($647 relative to $744), rather than 26% lower ($647 relative to $873) using our richest set of observables.[18]

Parametric Selection Model Finally, we consider a more parametric selection model as an alternative way to adjust for observables and unobservables. It yields a less aggressive correction for unobservables than the statistical correction shown in the last two rows of Table 9.

In this alternative approach, we continue to assume that mortality is not affected by enrollment in MA. We also assume that there is only one dimension of unobservable heterogeneity – which we can think of as health status – which affects both costs (under TM) and mortality. Under these assumptions, we can essentially use mortality rate as a "control function" for unobservables.

Specifically, we assume the data generating process arise from two equations. The first is a

[18] By way of comparison, Geruso and Layton (2015) estimate that MA risk scores are 6 to 16 percent higher than they would be for the same enrollee under TM. Taking their upper bound, such 16% "upcoding" alone would suggest that comparble TM spending would be $742 per enrollee month (16% lower than our baseline estimate of $884), which would imply that MA costs ($647) are13 percent lower than in TM. Put differently, the upper bound of their upcoding estimate could essentially account for all the selection on unboservables we estimate here.

selection equation

$$W_i = \mathbf{1}\{u\left(x_i, r_i, \eta_i\right) \leq 0\}, \tag{7}$$

where x_i is a vector of observables associated with beneficiary i, r_i is her risk score, and η_i is unobserved. The second equation is the potential outcome equation

$$E(\ln c_i^{TM}|x_i, r_i, \theta_i) = x_i'\alpha_x + \alpha_r(r_i + \theta_i). \tag{8}$$

We depart from the analysis thus far and consider the outcome to be log costs rather than dollars since it seems more natural to model factors (either observable or unobservable) as affecting cost proportionally.

Under these assumptions, the selection term (see equation (**??**)) is given by

$$S \equiv E\left[\ln c_i^{TM}|W_i = 1\right] - E\left[\ln c_i^{TM}|W_i = 0\right] = \alpha_r(E(\theta_i|x_i, r_i, W_i = 1) - E(\theta_i|x_i, r_i, W_i = 0)), \tag{9}$$

and the selection term is not zero when η_i and θ_i are correlated.

Suppose now that mortality realization is drawn from a logistic distribution, so that

$$E[m_i|x_i, r_i, \theta_i] = \frac{\exp\left(x_i'\beta_x + \beta_r\left(r_i + \theta_i\right)\right)}{1 + \exp\left(x_i'\beta_x + \beta_r\left(r_i + \theta_i\right)\right)},$$

where $m_i = 1$ if beneficiary (in either TM or MA) dies during the year (2010) and $m_i = 0$ otherwise. We can define now the log-odds ratio

$$\widetilde{m}_i(x_i, r_i, \theta_i) = \ln \frac{E[m_i|x_i, r_i, \theta_i]}{1 - E[m_i|x_i, r_i, \theta_i]} = x_i'\beta_x + \beta_r\left(r_i + \theta_i\right). \tag{10}$$

Equation (10) is estimable for the entire data, using the CMS data set, as we see mortality outcome there for both TM and MA populations. We can thus obtain unbiased estimates for β_x and β_r and back out $E(\theta_i|x_i, r_i, W_i)$. Equipped with this information, we can return to equation (8) and correct for the selection. That is, we can now regress $y_i'(x_i, r_i) = E(y_i^{TM}|x_i, r_i)$ on x_i and on $r_i + E(\theta_i|x_i, r_i, W_i = 0)$ and (given our assumptions) obtain unbiased estimates for α_x and α_r. We then have all the pieces to plug into the selection term, equation (9), and obtain the extent of selection on unobservables.

This approach yields that 23% of the residual difference in total spending between AM and (reweighted) TM is attributable to selection on unobservables. That is, we use equation (**??**), substituting $\ln c_i^{TM}$ for y_i^{TM} (using the reweighted cost) to compute the total difference in log cost and obtain 0.54. We then follow the above procedure to estimate S (see equation (9)), and we obtain 0.125, which is 23% of the overall difference.

8 Conclusion

We have compared healthcare spending and utilization in public and private Medicare. This setting provides a rare opportunity for a "side by side" comparison of public and private health insurance systems operating on a similar scale, for the same population, in the same markets, and with the same providers. Novel data from the Health Care Cost Institute on the healthcare claims of MA enrollees allowed us a rare look inside the "black box" of healthcare utilization and spending in MA.

We find that healthcare spending for enrollees in MA is 27% lower than for enrollees in TM in the same county and risk score. We estimate that 25-50 percent of this difference may reflect selection into MA on unobservable characteristics correlated with healthcare spending. The rest, presumably, reflects the "treatment effect" of MA. The lower spending by MA enrollees is entirely due to lower healthcare utilization. Prices appear similar in MA and TM. Where we can most directly measure this – the price of an admission for a given DRG at a given hospital – we estimate that average prices in MA are 1.5% higher than in TM. Reductions in utilization appear similar both for types of care where there is concern about "over use" (imaging and diagnostic tests) and where there is concern about "under use" (preventive care).

We provide suggestive evidence for some of the potential channels by which MA may reduce healthcare utilization for enrollees. We found that utilization is lower in MA but that, conditional on an encounter, spending per encounter is similar or slightly higher in MA. This suggests that MA manages to restrict utilization on the margin to sicker individuals. Relatedly, individuals discharged from the hospital are much more likely to be sent home – and less likely to be sent to post-acute care facility– if they are enrolled in MA rather than TM. We also find evidence consistent with substitution to less expensive types of care in MA: declines in specialist visits are much larger than declines in primary care visits, and while inpatient surgery rates are lower in MA, outpatient surgery rates are higher.

Finally, in light of the widespread interest in geographic variations in healthcare spending in TM, and recent work on geographic variations in commercial (under 65) private insurance, we explored similar comparisons in MA. Although geographic variation in spending in TM is often viewed as a reflection of the inefficiencies in a public health insurance system, we find similar – in fact slightly larger – geographic variation in spending in MA compared to TM. And while recent work has emphasized the much greater geographic pricing variation in private commercial insurance than in TM, we find similar – in fact slightly smaller – geographic variation in pricing in MA compared to TM.

One natural question these findings raise is their implications for MA insurers and consumers. For insurers, our estimates from MA data indicate that their revenue exceeds their healthcare expenditures by $180 per enrollee month. An important area for further work is to examine how this may be dissipated through other costs, such as the administrative costs of providing the insurance and the marketing costs of attracting enrollees. A related and important question is whether and how competitive pressures affect the MA market.

Implications for consumers are even more difficult to assess, since the elements of their objective function are not as straightforward to define or measure. A simple revealed preference argument would suggest that consumers who choose MA are better off in it. Other inferences are harder to make. Quality of the healthcare experience is difficult to assess; our measures of preventive care point to reductions there that are similar in magnitude to other forms of care. We can calculate the mean financial benefit (i.e. rebate to consumers as measured in the bid data) as $53 per enrollee-month, but, of course, financial benefits may be valued differently from their actuarial value, and MA plans have other attributes that will affect consumer surplus, such as limited networks. Further work on the implications of privately provided Medicare for both consumers and producers is an important area for further work.

References

Ayanian J.Z., Landon BE, Zaslovsky AM, Saunders R, Pawlson L.G., Newhouse JP. 2013. “Quality of Ambulatory Care in Medicare Advantage HMOs and Traditional Medicare.” *Health Affairs 32(7)*:1228-1235.

Baker, Laurence C., M. Kate Bundorf, Aileen M. Devlin and Daniel P. Kessler. 2016. "Medicare Advantage Plans Pay Hospitals Less than Traditional Medicare Pays." *Health Affairs* 35 (8): 1444-1451.

Brot-Goldberg, Zarek, Amitabh Chandra, Benjamin Handel and Jonathan Kolstad. 2015. “What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics. NBER Working Paper 21632.

Clemens, Jeffrey and Joshua Gottlieb. Forthcoming. “In the Shadow of a Giant: Medicare’s Influence on Private Physician Payments.” *Journal of Political Economy.*

Clemens, Jeffrey, Joshua Gottlieb and Timea Molnar. 2015. “The Anatomy of Physician Payments: Contracting Subject to Complexity.” NBER Working Paper 21642.

Cooper, Zack, Stuart Craig, Martin Gaynor, and John Van Reenen. 2015. “The Price Ain’t Right? Hospital Prices and Health Spending on the Privately Insured.” NBER Working Paper 21815.

Cutler, David, Mark McClellan, and Joseph P. Newhouse. 2000. “How does managed care do it?” *RAND Journal of Economics* 31(3): 526-548.

Curto, Vilsa, Liran Einav, Jonathan Levin and Jay Bhattacharya. 2014. “Can Health Insurance Competition Work? Evidence from Medicare Advantage.” NBER Working Paper 20818.

Duggan, Mark, Jonathan Gruber, and Boris Vabson. 2015. “The Efficiency Consequences of Health Care Privatization: Evidence from Medicare Advantage Exits.” NBER Working Paper 21650.

Einav, Liran, Amy Finkelstein and Maria Polyakova. 2016. “Private Provision of Social Insurance: Drug-specific Price Elasticities and Cost Sharing in Medicare Part D.” NBER Working Paper 22277.

Finkelstein Amy, Matthew Gentzkow and Heidi Williams. 2016. “Sources of Geographic Variation in Health Care: Evidence from Patient Migration.” *Quarterly Journal of Economics.*

Fisher, Elliott S., David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder. 2003a. “The Implications of Regional Variations in Medicare Spending: the

Content, Quality, and Accessibility of Care. Part 1." Annals of Internal Medicine, 138(4): 273–287.

Fisher, Elliott S., David Wennberg, Theresa Stukel, Daniel Gottlieb, F.L. Lucas, and Etoile L.Pinder. 2003b. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care" Annals of Internal Medicine 138(4): February 18, 2003. 288-299

Institute of Medicine, 2013. "Variation in health care spending: Target decision making, not geography." Washington DC: The National Academies Press.

Kaiser Family Foundation, 2010. https://kaiserfamilyfoundation.files.wordpress.com/2013/01/8080.pdf

Landon, Bruce, Alan Zaslavsky, Robert Saunders, L. Gregory Pawlson, Joseph Newhouse and John Ayanian. 2012. "Analysis of Medicare Advantage HMOs Compared with Traditional Medicare Shows Lower Use of Many Services During 2003-2009." *Health Affairs* No. 12: 2609-2617.

Geruso, Michael and Timothy Layton. 2015. "Upcoding: Evidence from Medicare on Squishy Risk Adjustment." NBER Working Paper 21222.

McGuire, Thomas, Joseph P. Newhouse, and Anna D. Sinaiko. 2011. "An Economic History of Medicare Part C." *The Milbank Quarterly* 89(2): 289-332.

Philipson, Tomas, Dana Goldman, Seth Seabury, Darius Lakdawalla and Lee Lockwood. 2010. "Geographic Variation in Health Care: The Role of Private Markets." *Brookings Papers on Economic Activity,* Spring, pages 325-355.

Skinner, Jonathan. 2011. "Causes and Consequences of Regional Variations in Health Care." In Handbook of Health Economics. Vol. 2, eds. Mark V. Pauly, Thomas G. McGuire and Pedro P.Barros, 45–93. Elsevier.

Song, Yunjie, Jonathan Skinner, Julie Bynam, Jason Sutherland, and Elliott Fisher. 2010. "Regional Variations in Diagnostic Practices." New England Journal of Medicine, 363(1): 45–53.

Newhouse, Joseph and Thomas McGuire. 2014. "How Successful in Medicare Advantage?" *The Milbank Quarterly* 92(2): 351-394.

U.S. Government Accountability Office, 2008. Medicare Part B Imaging Services: Rapid Spending Growth and Shift to Physician Offices Indicate Need for CMS to Consider Additional Management Practices. URL http://www.gao.gov/products/GAO-08-452 (accessed 2.23.15)

Figure 1: MA penetration over time

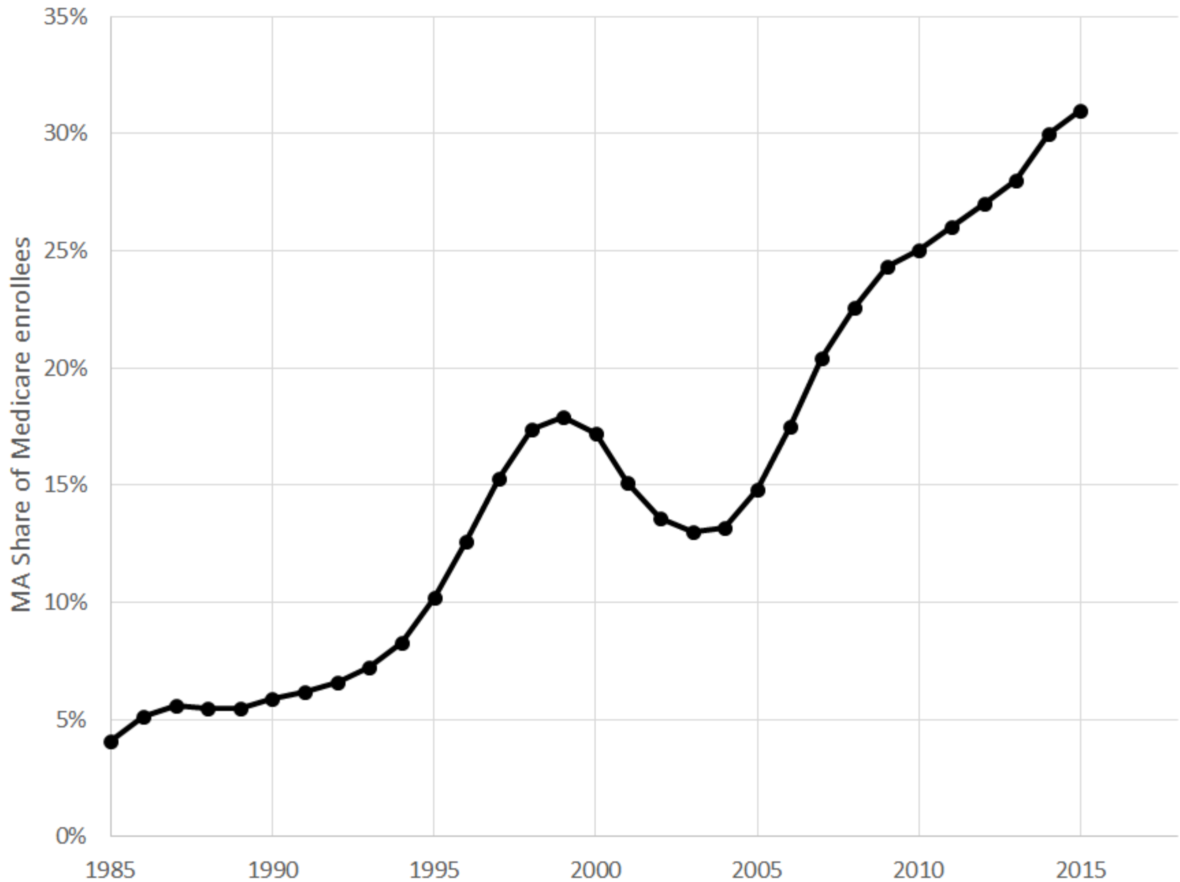

Figure shows the share of Medicare beneficiaries enrolled in Medicare Advantage plans, year by year. The data source is CMS's Medicare Managed Care Contract Plans Monthly Summary Reports. All data are from December of the year indicated.

Figure 2: State-by-State Comparison of TM and MA Spending

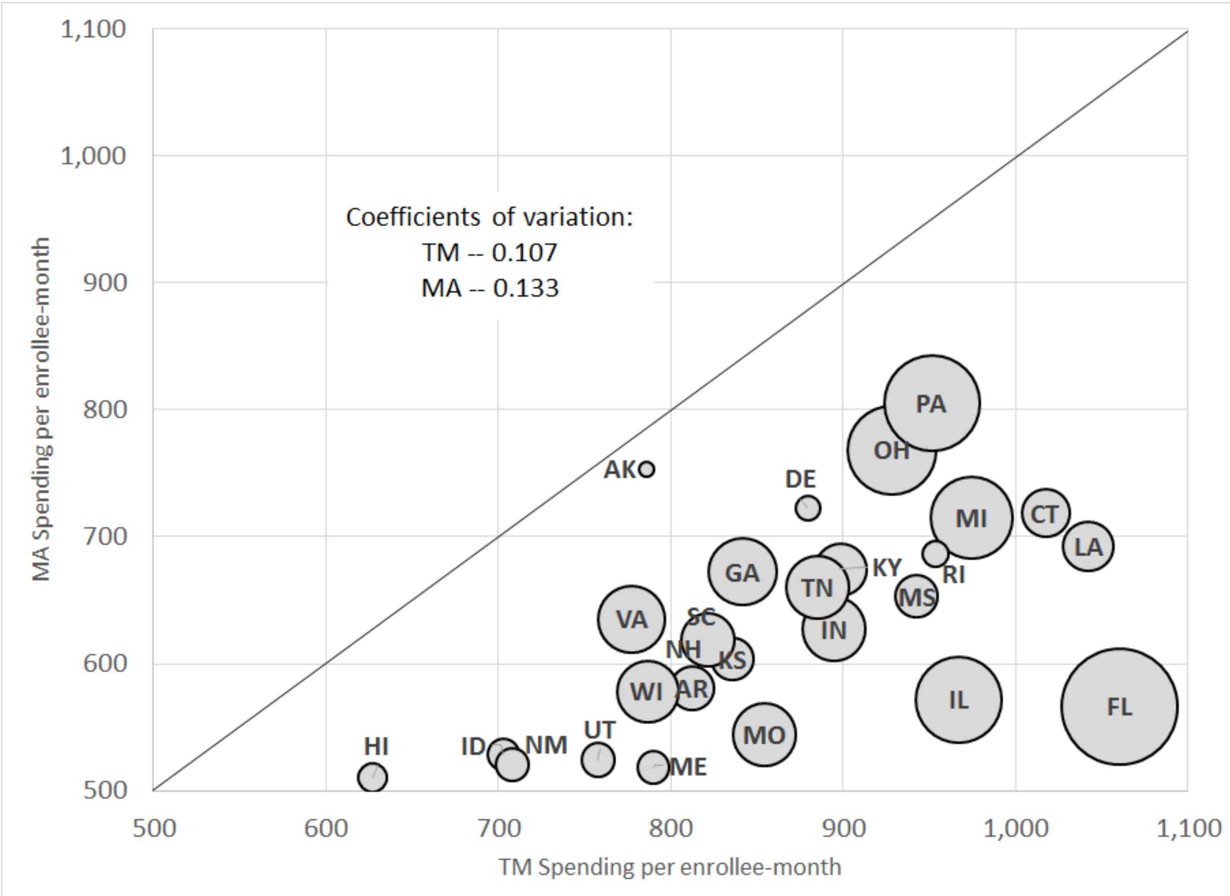

Figure plots MA spending per enrollee-month against TM spending per enrollee-month for our baseline sample (see Table 1, columns (8) and (6)) , for each of the 27 states in our baseline sample. The size of each bubble is proportional to the number of total Medicare enrollees in the state.

Figure 3: TM-MA Spending Differences across States

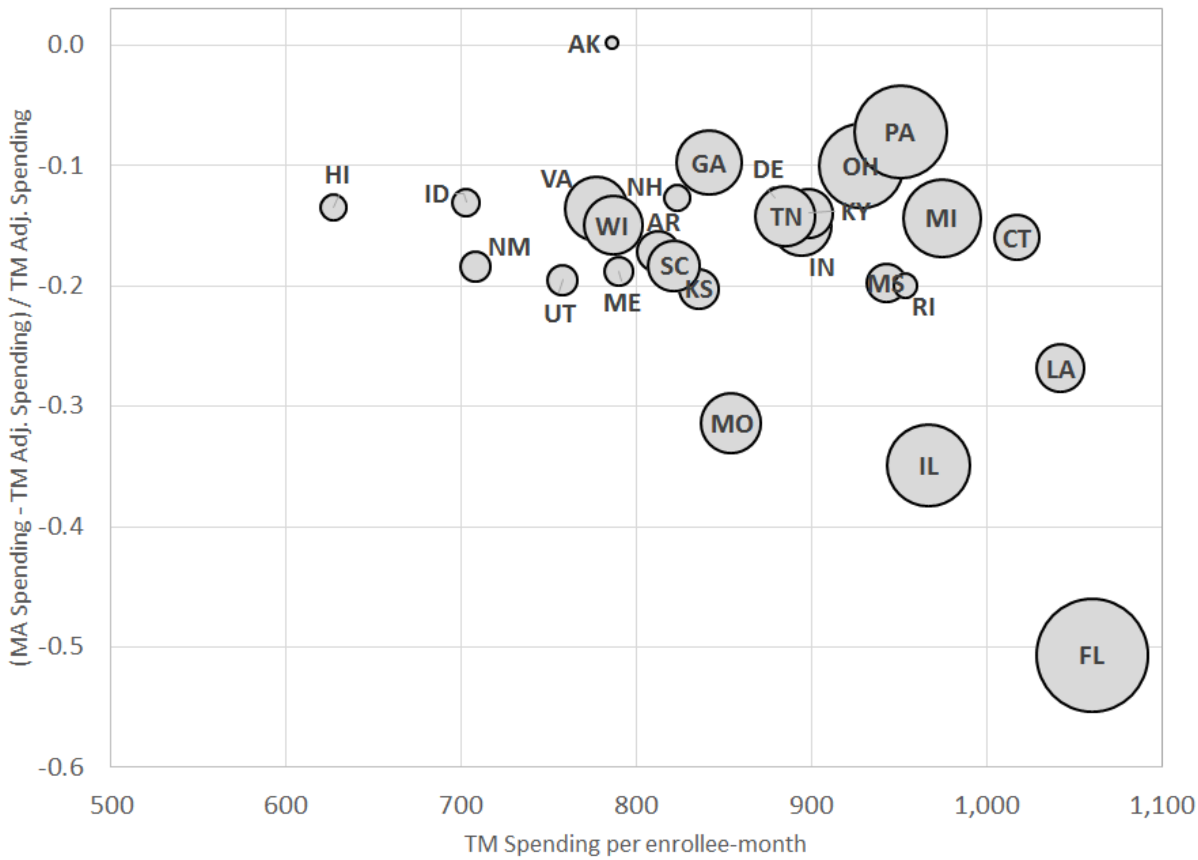

Figure plots the (percentage) difference between average MA spending and (re-weighted) TM spending per enrollee-month against average TM spending for our baseline sample (see Table 1, columns (8) and (6)), for each of the 27 states in our baseline sample. The size of each bubble is proportional to the number of total Medicare enrollees in the state. The y-axis compares MA spending to TM spending that is re-weighted to match the MA population on county and risk score, using our preferred weighting (see Table 2, column (4)). The x-axis reports average (unadjusted) TM spending in the state (see Table 2, column (1)).

Figure 4: TM-MA price differences for inpatient admissions, across DRGs

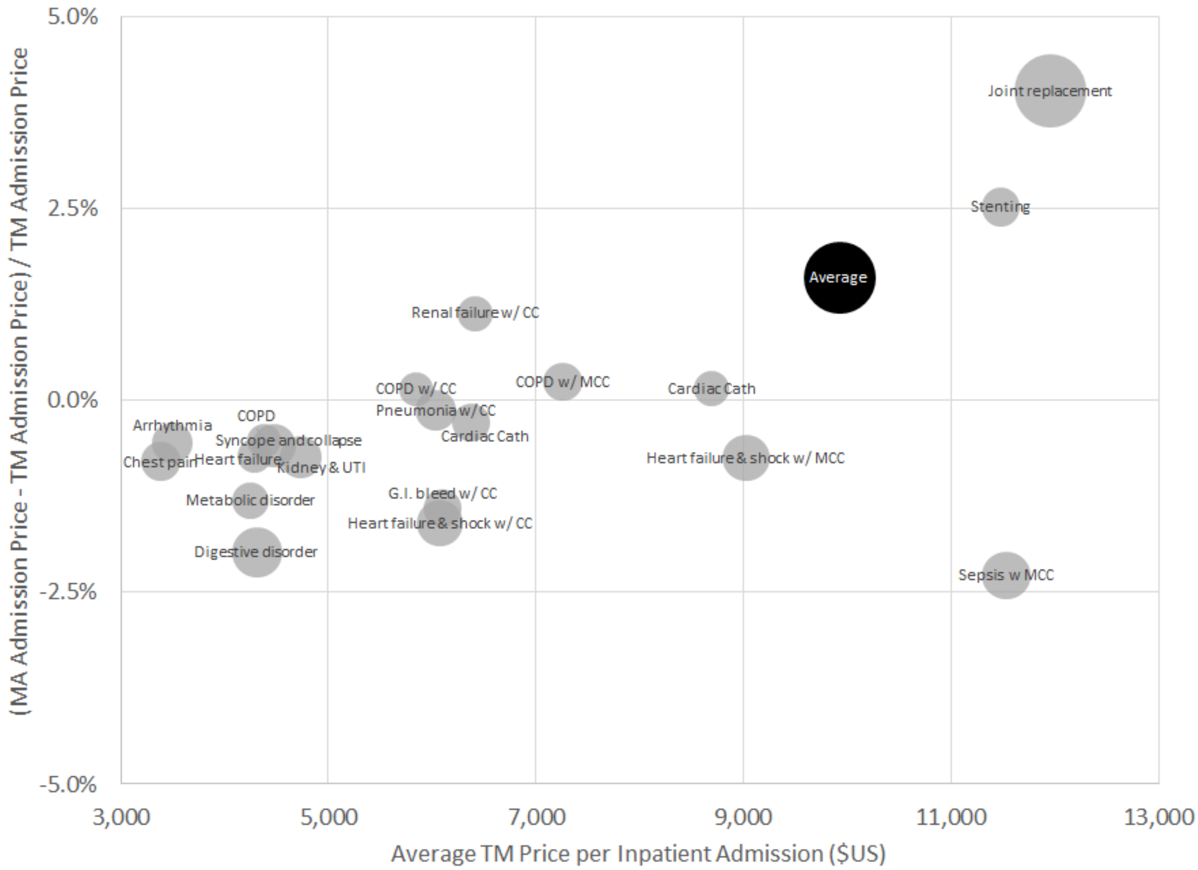

Figure plots the (percentage) difference between average MA prices and TM prices for a hospital admission, overall and for the 20 most common DRGs. Averages are computed for each DRG using a common (MA) "basket" of state admission shares. Sample is a subset of our baseline sample; it is limited to all MA inpatient admissions to hospitals that are paid (by CMS) under prospective payment system (PPS). The figure shows results for the top 20 DRGs and the average across all DRGs, not only the top 20. The size of each bubble (except for the overall "Average" bubble) is proportional to the number of MA admissions with that DRG.

Figure 5: TM-MA price differences for inpatient admissions, across states

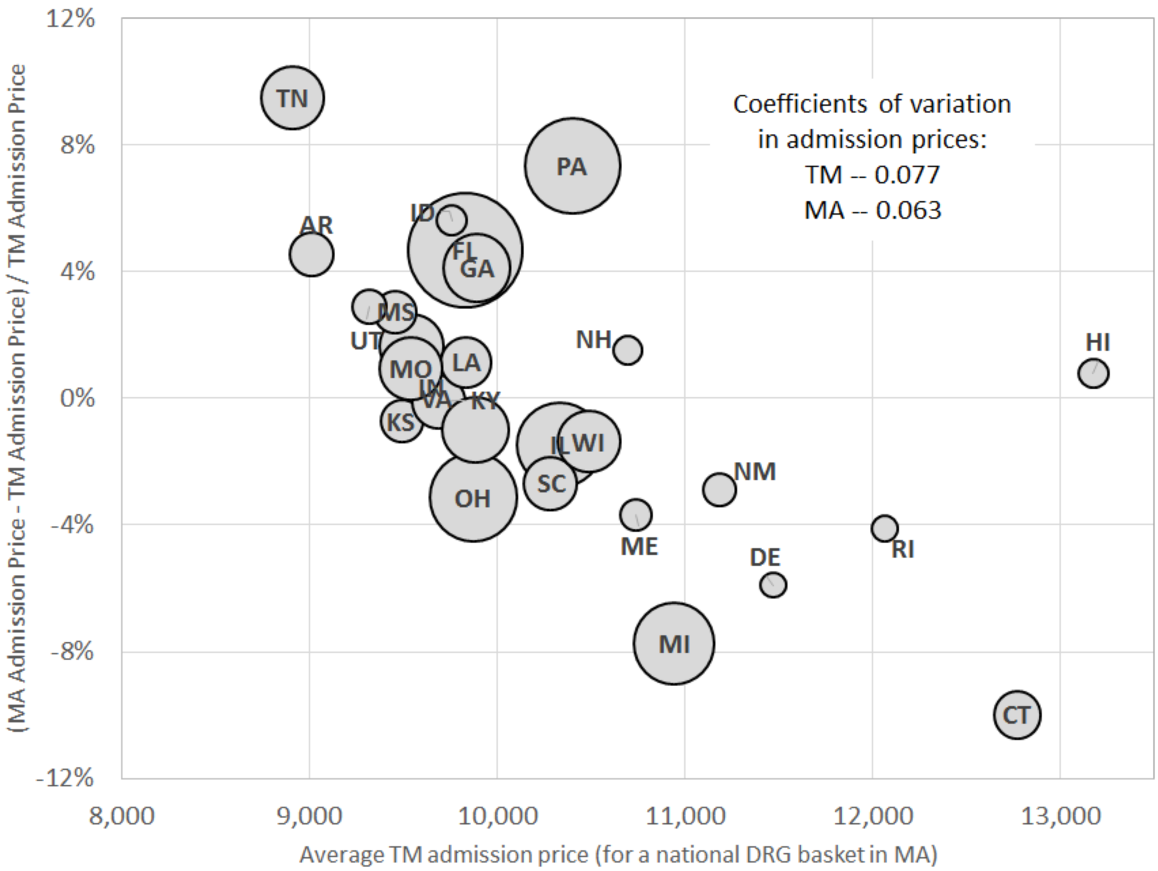

Figure plots the (percentage) difference between average MA prices and TM prices for a hospital admission for each state. Averages are computed for each state using a common (MA) "basket" of DRG admission shares. Sample is a subset of our baseline sample; it is limited to all MA inpatient admissions to hospitals that are paid (by CMS) under prospective payment system (PPS). The size of each bubble is proportional to the number of MA admissions in that state. Data for Alaska is omitted because it had too few inpatient admissions for us to report. We also estimate a coefficient of variation across states in MA prices for inpatient admissions of 0.063, compared to 0.077 in TM; each statistic is computed using total Medicare enrollees in the state as weight.

Table 1: Baseline sample

Data source / sample	All CMS[a]			Baseline CMS[b]			All HCCI[a]	Baseline HCCI[b]
	TM	MA (all insurers)	MA (HCCI insurers)	TM	MA (all insurers)	MA (HCCI insurers)	MA (HCCI insurers)	MA (HCCI insurers)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Enrollee-level summary[e]								
No. of enrollees (000s)	27,066	10,502	3,904	13,577	4,563	2,016	2,933	2,039
Female	0.574	0.574	0.574	0.577	0.569	0.571	0.569	0.573
Age[c]	75.1	74.3	74.2	75.1	74.2	73.9	--	--
Coarse age:[c,d]								
65-74	0.534	0.569	0.573	0.530	0.575	0.591	0.592	0.588
75-84	0.320	0.318	0.315	0.323	0.319	0.307	0.306	0.309
85+	0.146	0.113	0.111	0.147	0.107	0.103	0.102	0.103
Dual eligible[c]	0.141	0.122	0.110	0.131	0.074	0.071	--	--
SNP enrollees[c]	--	0.080	0.064	--	0.000	0.000	0.000	0.000
Risk score	1.074	1.015	1.021	1.082	1.022	1.023	--	--
Died in 2010	0.049	0.036	0.036	0.051	0.036	0.035	--	--
Panel B: Spending per enrollee-month[f]								
No. of enrollee-months (000s)	306,801	117,423	43,738	154,082	51,299	22,649	32,423	22,609
Total Spending ($)	933	--	--	924	--	--	639	647
Insurer Spending ($)	793	--	--	783	--	--	586	596
OOP Spending ($)[g]	139	--	--	141	--	--	53	52
Panel C: Payments to insurers per enrollee-month[f]								
Overall CMS expenditure ($)[h]	--	818	818	--	778	787	--	--
Actuarial value of incremental consumer benefits ($)[i]	--	63	53	--	58	53	--	--
Plan payments for organic MA services ($)[j]	--	799	805	--	761	776	--	--

Table presents summary statistics for various sample definitions. Our baseline sample is summarized in columns (6) and (8), highlighted in gray.

[a] Sample include all Medicare enrollees who are 65 or older by the end of 2010.

[b] Baseline sample excludes SNP enrollees, enrollees in masked (i.e. very small) zip codes, and enrollees in states in which the number of enrollee-months in HCCI is not within 10% of that in CMS. See Appendix.

[c] These variables are defined as of the first month in which we observe the enrollee during 2010 (January in the vast majority of cases).

[d] In HCCI we only have information about age in three bins: 65-74, 75-84, and 85+.

[e] At the enrollee-level, we define an individual as enrolled in TM if she is never enrolled in MA during the sample year and is enrolled in TM for at least one month of the sample year; we define her as enrolled in MA if she is enrolled in MA in any month of the year, and we assign her to an HCCI insurer if she is covered by one of them in her first month in MA. Dual eligibility and SNP enrollment is likewise defined based on the first month in which an enrollee is observed during the sample year.

[f] We count an enrollee-month in TM if she is enrolled in TM that month and never enrolled in MA during the sample year; any enrollee-months in MA (or in HCCI insurers) are counted as such.

[g] Out of pocket (OOP) spending denotes amount owed by enrollee. For TM enrollees, OOP Spending may be partially covered by supplemental (Medigap or employer-sponsored) coverage.

[h] This includes all payments made from CMS to the MA plans, including risk-adjusted payments and rebates.

[i] This is also known as the "rebate."

[j] The variable "Plan payments for organic MA services ($)" is equal to "Overall CMS expenditure ($)" plus additional premiums paid by the beneficiaries minus the non-cost-sharing component of the rebate.

Table 2: Baseline reweighting

Source	CMS				HCCI
Sample	TM	TM	TM	TM	MA
Reweight by	None	County	County & Risk bin 0.5	County & Risk bin 0.1	None
	(1)	(2)	(3)	(4)	(5)
No. of enrollee-months (000s)	154,082	154,082	154,082	154,082	22,609
Total Spending ($/month)	924	950	889	884	647
Insurer Spending ($/month)	783	806	753	749	596
OOP Spending ($/month)[a]	141	144	135	135	52

Results based on baseline sample (see Table 1, columns (8) and (6)). All statistics are at the enrollee-month level.

[a] Out of pocket (OOP) spending denotes amount owed by enrollee. For TM enrollees, OOP Spending may be partially covered by supplemental (Medigap or employer-sponsored) coverage.

Table 3: Spending differences for different groups of enrollees

	% MA enrollees	TM, unweighted	TM, weighted[a]	MA	Difference	
					(4)-(3)	((4)-(3)) / (3)
	(1)	(2)	(3)	(4)	(5)	(6)
No. of enrollee-months (000s)	22,609	154,082	154,082	22,609		
Total Spending	100%	924	884	647	-237	-26.8%
Panel A. Spending ($/month) by enrollee characteristics						
Male	42.7%	928	887	681	-206	-23.2%
Female	57.3%	922	882	623	-260	-29.4%
65-74	56.1%	729	679	544	-134	-19.8%
75-84	32.7%	1,041	1,009	737	-272	-27.0%
85+	11.2%	1,299	1,312	904	-409	-31.2%
Urban[b]	80.7%	952	918	648	-270	-29.4%
Rural[b]	19.3%	861	756	646	-110	-14.6%
Panel B. Realized distribution of spending ($/month)						
% w/ no spending		0.37	0.38	0.46	0.08	21.6%
Median spending		95	90	39	-51	-56.6%
75th pctile		338	330	224	-106	-32.1%
90th pctile		1,339	1,291	859	-432	-33.4%
95th pctile		3,494	3,290	2,179	-1,112	-33.8%
97.5th pctile		8,454	7,863	5,731	-2,132	-27.1%
99th pctile		18,759	17,830	13,693	-4,137	-23.2%

Results based on baseline sample (See Table 1, columns 8 and 6). All statistics are at the enrollee-month level. All spending numbers are in $/month.

[a] Weighting based on our preferred weighting, as in column (4) in Table 2.

[b] Rural/urban assignment is based on enrollee zip code. A zip code is defined as rural if it does not belong to an MSA

Table 4: Spending differences for different components of spending

	TM, unweighted[a]	TM, weighted[a]	MA	Difference	
				(3)-(2)	((3)-(2)) / (2)
	(1)	(2)	(3)	(4)	(5)
No. of enrollee-months (000s)	154,082	154,082	22,609		
Total spending[b]	924	884	647	-237	-26.8%
Inpatient	370	344	272	-72	-21.0%
Outpatient	458	450	331	-119	-26.5%
SNF	97	91	45	-45	-50.1%
Hospice[c]	31	33	24	-9	-27.7%

Results based on baseline sample (See Table 1, columns 8 and 6). All statistics are at the enrollee-month level. All spending numbers are in $/month.

[a] Weighting based on our preferred weighting, as in column (4) in Table 2.

[b] Total spending is the sum of inpatient, outpatient, and SNF spending. It doesn't include hospice.

[c] Hospice expenditures for MA enrollees are billed directly to CMS, so for MA enrollees they are in fact observed in the CMS data and not in the HCCI data.

Table 5: Geographic variation across components of spending

	Corr (TM,MA) (1)	Coeff. Of Var (TM) (2)	Coeff. Of Var (MA) (3)
Total spending	0.207	0.107	0.133
Outpatient spending	-0.197	0.134	0.141
SNF spending	0.623	0.184	0.335
Inpatient spending	0.724	0.111	0.140

Results based on baseline sample (See Table 1, columns 8 and 6). Total spending and its components are defined (at the enrollee-month level) as in Table 4. Table reports correlations and coefficients of variation in spending per enrollee-month across states; all these statistics are weighted, using total Medicare enrollees in the state as weights. TM spending is unadjusted for demographics (see Table 2, column 1).

Table 6: Differences in healthcare utilization and in spending per encounter

	TM, unweighted	TM, weighted[a]	MA	Difference	
				(3)-(2)	((3)-(2)) / (2)
	(1)	(2)	(3)	(4)	(5)
Total spending ($/month)	924	884	647	-237	-26.8%
Panel A. Utilization measures (per enrollee-month)					
Inpatient days	0.203	0.187	0.145	-0.043	-22.8%
Any inpatient admission	0.027	0.025	0.021	-0.0045	-17.8%
Days cond'l on any	7.44	7.41	6.96	-0.45	-6.1%
SNF days	0.336	0.308	0.132	-0.176	-57.2%
Days cond'l on any	47.4	46.8	20.7	-26.1	-55.8%
Outpatient ED visits	0.030	0.027	0.023	-0.005	-16.5%
Physician visits	1.24	1.24	1.02	-0.22	-18.1%
Any physician visits	0.549	0.548	0.488	-0.060	-11.0%
Number of visits cond'l on any	2.25	2.27	2.09	-0.18	-8.0%
Panel B. Spending per encounter ($)					
Spending per SNF day	377	376	384	8	2.0%
Spending per outpatient ED visit	783	773	846	74	9.6%
Inpatient:[b]					
Spending per admission	10,223	10,201	10,148	-53	-0.5%
Spending per day	1,906	1,900	1,912	12	0.6%
Spending per AMI admission	14,619	14,580	14,845	266	1.8%
Spending per AMI day	2,725	2,726	2,661	-65	-2.4%

Results based on baseline sample (See Table 1, columns 8 and 6). All statistics are at the enrollee-month level, but all expenditures or days associated with a given encounter are attributed to the original admission date, even if it extends beyond the month.

[a] Weighting based on our preferred weighting, as in column (4) in Table 2.

[b] Inpatient spending here includes only payments to the hospital, it does not include associated physician payments as in prior tables.

Table 7: Utilization differences across different types of care

	TM, unweighted	TM, weighted[a]	MA	Difference	
				(3)-(2)	((3)-(2)) / (2)
	(1)	(2)	(3)	(4)	(5)
A. Testing and imaging:					
Diagnostic tests	2.14	2.10	1.54	-0.56	-26.6%
Any diagnostic test	0.356	0.347	0.293	-0.055	-15.8%
Cond'l on any	6.00	6.05	5.28	-0.77	-12.8%
Imaging procedures	0.67	0.65	0.52	-0.13	-20.3%
Any imaging test	0.177	0.175	0.155	-0.020	-11.3%
Cond'l on any	3.76	3.73	3.35	-0.38	-10.1%
B. Preventive care (rates per relevant population):[b]					
Flu shot	0.051	0.050	0.031	-0.019	-38.0%
Cardiovascular screen	0.092	0.095	0.077	-0.017	-18.4%
Colorectal cancer screen	0.010	0.010	0.009	-0.002	-15.2%
Mammogram	0.046	0.046	0.047	0.002	3.3%
Pap smear	0.012	0.012	0.013	0.001	8.6%
Prostate cancer screen	0.024	0.023	0.018	-0.006	-24.1%
Hemoglobin A1c test	0.065	0.064	0.055	-0.009	-14.1%
Blood lipids test	0.105	0.108	0.091	-0.017	-15.9%
Eye exam	0.068	0.068	0.054	-0.014	-21.1%

Results based on baseline sample (See Table 1, columns (8) and (6)). All statistics are at the enrollee-month level.

[a] Weighting based on our preferred weighting, as in column (4) in Table 2.

[b] Rates are per the relevant population, which is: everyone for flu shot, cardiovascular screen, and colorectal cancer screen; women for pap smear; women aged 65-74 for mammogram; men for prostate cancer screen; and enrollees aged 65-74 with a diabetes diagnosis for hemoglobin test, blood lipids test, and eye exam.

Table 8: Potential channels for cost saving

	TM, unweighted	TM, weighted[a]	MA	Difference	
				(3)-(2)	((3)-(2)) / (2)
	(1)	(2)	(3)	(4)	(5)
A. Hospital discharge destinations:					
Home	0.0125	0.0123	0.0109	-0.0014	-11.8%
Home health service org.	0.0048	0.0049	0.0039	-0.0010	-21.0%
SNF	0.0060	0.0062	0.0039	-0.0023	-36.6%
Other post-acute care	0.0009	0.0009	0.0004	-0.0005	-57.3%
Other (incl. hospice, death)	0.0031	0.0030	0.0018	-0.0013	-41.9%
B. Surgeries and specialists:					
Total surgeries	0.036	0.033	0.039	0.006	18.3%
Outpatient surgeries	0.028	0.025	0.032	0.007	26.1%
Inpatient surgeries	0.008	0.008	0.007	-0.001	-8.0%
Primary care visits	0.381	0.376	0.352	-0.024	-6.3%
Specialist visits	0.855	0.866	0.665	-0.201	-23.2%

Results based on baseline sample (See Table 1, columns (8) and (6)). All statistics are at the enrollee-month level. All spending numbers are in $/month. Panel A reports (unconditional) hospital discharge destinations.

[a] Weighting based on our preferred weighting, as in column (4) in Table 2.

Table 9: Alternative ways to correct for selection into MA

Reweight on	Covariates	TM Mean Total Spending (Reweighted)	
		Reweight nationally	Reweight county-by-county
(1)	(2)	(3)	(4)
1. None		924	924
2. Risk score[a]	--	879	884
3. Prop. score	county*risk score[a]	886	887
4. Prop. score	county*(age FE, female, HCC FE)[b]	894	894
5. Prop. score	county*(age FE, female, HCC FE, dual)[c]	873	874
6. Prop. score	county*dual*(age FE, female, HCC FE)[d]	870	870
7. Predicted mortality[e]		746	744
8. Prop. score	county*predicted mortality[e]	748	747

Results based on baseline sample (See Table 1, columns (8) and (6)). The "propensity score" approach in rows 4-6 and row 8 is based on a logistic regression (estimated separately, county by county) for being in MA, using the covariates listed in Column (2). Rows 2 and 7 use our baseline re-weighting approach (see equation (2)) with the re-weighting based on risk score bin (row 2) or predicted mortality bin (row 7).

[a] Risk scores are mapped to 0.1 bins, and are included using indicator variables for each bin.

[b] The independent variables are dummies for age, gender, and the 70 hierarchical condition categories (HCCs) that appear in the MA risk adjustment model.

[c] The independent variables are dummies for age, gender, the 70 hierarchical condition categories (HCCs) that appear in the MA risk adjustment model, and dual eligibility for Medicare and Medicaid.

[d] The independent variables are dummies for age, gender, and the 70 hierarchical condition categories (HCCs) that appear in the MA risk adjustment model. Each of these dummies is interacted with a dummy for dual eligibility for Medicare and Medicaid.

[e] "Predicted mortality" is generated based on a regression of a annual mortality indicator on indicators for risk bins of 0.1. The regression is run separately for MA enrollees and for TM enrollees. The resultant mortality prediction is included in mapped to bins of 0.001, and included as indicator variables for each bin.

Appendix A: Construction of the baseline sample

A.1 Raw data files

HCCI Files We have data from HCCI on a convenience sample of 2010 Medicare Advantage (MA) enrollees in three insurers: Aetna, Humana, and UnitedHealthCare (hereafter, "HCCI insurers"). The data were provided to HCCI by the private insurers and exclude enrollees in highly capitated plans, Special Needs Plans, plans with various data issues, and other limitations.[19]

The HCCI data contain four main files. There is an enrollment file, which we use to define the sample and obtain basic demographic information. The unit of observation is an enrollee-month. The enrollment file contains monthly indicators for enrollment, age (in bins of 10 years), gender, the enrollee's state of residence, and the enrollee zip code (masked for zip codes with a 2010 census population of less than 1,350). We observe "exit" within the year from the HCCI data but do not directly measure mortality. The data do not contain indicators for which insurer (or plan) the enrollee is covered by. In addition, there are three claims files – inpatient, outpatient, and physician – which we use to measure medical spending. In these files the unit of observation is a claim, payable by one of the HCCI insurers to a medical provider.

CMS Files We have data from CMS on the universe of individuals enrolled in Medicare at any point in 2010. This includes both those enrolled in Traditional Medicare (TM) and those enrolled in MA. For all enrollees – both those in TM and those in MA – we have four main files: the enrollment data base (EDB), the common Medicare enrollment file (CME), the Health Plan Management System (HPMS), and the Risk Adjustment Processing System (RAPS). The two enrollment files allow us to observe for every enrollee: exact date of birth, date of death (if applicable), gender, and zip code. They also include monthly data on whether the individual is enrolled in TM Part A, enrolled in TM Part B, enrolled in MA, whether they are dually covered by Medicare and Medicaid, and whether the individual died; note that dual coverage and mortality are observed in the CMS files for both MA and TM enrollees.

For enrollee-months in MA we also observe a plan identifier. Using the HPMS plan-level data on the parent organization, we are able to identify which plans are provided by the HCCI insurers, and also whether the plan is a Special Needs Plan (SNP), specialized Medicare Advantage plans for particular types of individuals (e.g. those in long term care institutions). We assign an MA enrollee an MA plan based on the first plan in which she is enrolled in the year.

The RAPS file has a risk score and indicators for each health indicator (HCC) that goes into the calculation of the risk score, for every enrollee. These HCCs are then integrated using a predictive formula that combines them together to form a risk score, which is a predictor of the enrollee's healthcare spending in the subsequent year. We observe these indicator for MA enrollees since MA plans must submit HCCs to CMS to determine their CMS payments. The RAPS file also contains

[19] The description of the exclusion critiera come from HCCI, except for the exclusion of SNPs which we determined by looking at the type of plan codes that apopear in the HCCI enrollment file.

indicators for the enrollees' type – community (90%), new (9%), or long-term institutional (1%) – and three risk scores (one for each type), and we assign each enrollee her type-specific risk score.

For TM enrollees only, the CMS data allows us to measure healthcare utilization and spending through 6 claims files: inpatient, outpatient, SNF, home health, durable medical equipment, and physician. A seventh claims file – the hospice claims file – contains utilization and spending for both TM and MA enrollees (since hospice is reimbursed by CMS for MA enrollees as well as TM enrollees); the hospice file is the only CMS file where we can observe utilization and spending for MA enrollees.

Finally, for MA enrollees we use the Monthly Membership Detail Report and the HPMS to construct information on revenues to MA insurers. Specifically, for each individual enrolled in an MA plan, we observe the payment from CMS to the insurer. The payment from CMS to the insurer consists of a part that is retained by the insurer and the rebate which is passed on by the insurer to the enrollee. We observe, for each plan, this rebate amount, as well as the Part C premium that is paid by the enrollee to the insurer. We define MA revenue for a given enrollee-month as the payments from CMS to the insurer minus the rebate to consumers, plus the Part C premiums.

A.2 Sample definition

We use the HCCI data to analyze spending and healthcare utilization for individuals covered by the HCCI insurers. We use the CMS data for two primary purposes: to construct comparison spending and healthcare utilization estimates for "comparable" TM enrollees, and to create an independent measure of enrollment in the HCCI insurers' plans that we use to examine and validate the completeness of the HCCI enrollment data. Both of these exercises require that we define a TM and an MA enrollee in the CMS data.

Throughout this paper, in the CMS data we define an enrollee as enrolled in MA if she is enrolled in MA for at least one month during 2010; we define someone as enrolled in TM if she is not enrolled in MA during any month in 2010, and is enrolled in TM Part A and TM Part B in at least one month during 2010. We count the enrollee-months in MA as the total number of months in MA during the year. Within MA, we can further identify the subset of MA enrollees who are in the three HCCI insurers. We restrict our analysis to enrollee-months who are 65 and over, who reside in one of the 50 states or the District of Columbia; we do not require individuals to be enrolled for a full year.

We can measure the completeness of the HCCI data in terms of enrollment by the HCCI insurers by comparing enrollee-month counts in the HCCI data to enrollee-month counts for these HCCI insurers in the CMS data, which in principle records the universe of enrollees in those same plans. Appendix Table 1 shows enrollee-month counts for the three HCCI insurers according to the HCCI data and the CMS data, overall, and separately by state. For this analysis we exclude enrollees in HCCI who are in masked zip codes, and correspondingly exclude any individuals in the CMS data who are enrolled by an HCCI insurer in a zip code that does not appear in the HCCI data. We also exclude from the CMS enrollment counts any enrollees in SNP plans since, as discussed, these are also excluded from the HCCI data..The HCCI data contain about 80 percent of total MA enrollees for the HCCI insurers; "missing" enrollees disproportionately concentrated in the Western US.

We restrict our analysis to the 27 "complete data" states, which we define as states where the count of enrollee-months in HCCI is within 10 percent of the corresponding count in CMS data. The 10 percent cutoff is arbitrary, but 20 of the 27 states are within 5 percent, and these 20 states would account for more than 90% of the enrollees in the baseline sample, so the results are unlikely to change much with more conservative sample definitions. Using CMS data, Appendix Table 1 shows, by state, the MA share of Medicare enrollment and the HCCI insurer share of MA. Overall, the 27 states that we analyze comprise 57 percent of enrollment in HCCI insurers nation-wide.

Appendix B: Construction of specific variables

We analyze MA medical spending and utilization in the HCCI data. We benchmark it against TM spending and utilization in the CMS data, for observably similar enrollees. We therefore construct parallel medical spending and healthcare utilization variables in the HCCI and CMS data. Unless explicitly noted, all MA medical spending and healthcare utilization measures are derived from HCCI data, and all TM spending measures are derived from CMS data. All measures are constructed at the enrollee-month level unless explicitly noted.

Total spending is defined as the sum of insurer spending plus out-of-pocket spending. **Insurer spending** is defined based on the actual amount paid by the plan (either MA or TM) to the provider. In other words, it is the transacted (as opposed to list) price. **Out-of-pocket spending** is the amount owed by the enrollee (i.e. the sum of any coinsurance, copay, and deductible). For individuals enrolled in TM, some of this "out of pocket" spending may be covered by supplemental private insurance (Medigap), which they may purchase separately.

Medical spending is divided across claims files based on who is billed, which does not map perfectly to our concept of "place of care." In particular, institutional billing goes to the relevant institutional file (e.g., inpatient or outpatient) while individual provider billing (regardless of whether it is inpatient or outpatient) goes to the physician (aka carrier) file. The structure of claims files is slightly different across the two data sources. We use three HCCI claims files: Inpatient, outpatient and physician. We use seven CMS claims files: inpatient, outpatient, physician, SNF, home health, durable medical equipment, and hospice. In HCCI, the SNF spending is in the inpatient file; we identify SNF claims in the HCCI inpatient file based on their Place of Service (POS) codes (POS code of 31-33 determines a SNF). In HCCI, home health and durable medical equipment are in the outpatient and physician files. Hospice is reimbursed by TM for both TM and MA enrollees; there is therefore no hospice spending in the HCCI data, but we can observe hospice spending in the CMS data for both TM and MA enrollees. Finally, we note that in HCCI the inpatient file includes all admissions in 2010, while in CMS the inpatient and SNF files include discharges in 2010; we therefore supplement the 2010 SNF and inpatient discharge files in CMS with the 2011 SNF and inpatient discharge files, and in both files limit the analysis to admissions that occur in 2010; in this way we reconstruct a 2010 admission file that is parallel to the HCCI admission file.

Below we describe he construction of specific variables.

Total spending and components All of these measures are constructed at the enrollee-month level unless explicitly noted otherwise. Note that for inpatient and SNF spending, we associate the spending with the month in which the admission occurred even when the stay extends into subsequent months.

- **Total spending:** the sum of inpatient, outpatient, and SNF spending.
- **Inpatient spending:** in the CMS data it covers all spending on the inpatient file plus spending on the physician file associated with an inpatient hospital (POS code of 21). In the HCCI data it covers all spending on the inpatient file minus SNF spending (as mentioned, POS codes of 31-33) plus spending on the physician file associated with an inpatient hospital (POS code of 21).
- **Outpatient spending**: in CMS data it is the sum of all spending on the outpatient file, the home health file, and the durable medical equipment file, plus all spending on the physician file for which POS is not 21. In HCCI data is it the sum of all spending on the outpatient file (which, recall, includes home health and durable medical equipment), plus spending on the physician file for which POS is not 21.
- **SNF spending:** in CMS data it is the sum of all spending on the SNF file, while in HCCI file it is the sum of all spending on the inpatient file with POS codes 31-33.
- **Hospice spending**: hospice care is reimbursed by TM for both TM and MA enrollees. There is therefore no hospice spending in the HCCI data, but we can observe hospice spending in the CMS data for both TM and MA enrollees. We use the hospice file in the CMS data to measure hospice spending in TM and in MA.

Healthcare utilization In addition to measuring spending, we also measure healthcare utilization. We define a number of standard measures of healthcare use for each enrollee-month. We measure inpatient utilization using the inpatient files. In the HCCI data we only count observations that are inpatient hospital admissions (i.e. we exclude SNF admissions based on POS codes of 31-33). We measure SNF utilization using the SNF file in the CMS data and the inpatient file in the HCCI data, only counting admissions with POS codes of 31-33.

- **Inpatient days:** the sum of the days associated with each inpatient admission that month; as with our inpatient spending measure, this will include all the days for each admission in a given month, even if those days extend beyond that month. We measure the days of a given admission as the difference between discharge date and admission date, plus 1.

- **SNF days:** is defined analogously to inpatient days. In the CMS file, discharge date is missing for about 18 percent of the observations, which appears to reflect discharges that extend beyond the 100-day coverage period for SNF in TM. Since we are interested in TM-covered utilization, we impute 100 days for such discharges.

- **Inpatient admissions:** any inpatient admission that month.[20]

- **Physician visits:** is measured based on claims in the physician file (excluding claims with POS code of 21, which indicates that they occur in an inpatient setting). We define physician visits as the sum of **primary care visits** and **specialty care visits**. We allow a maximum of one primary care visit per patient-day, and one specialist visit per patient day. Following the approach in Finkelstein et al. (2016), our definition of primary care physicians and specialists follows the Dartmouth Atlas.[21] Specifically, we crosswalk the primary care and specialist definitions in the Dartmouth Atlas to the list of HCFA specialty codes in the CMS data. The HCCI data has a separate set of provider category codes which we crosswalk to the HCFA specialty codes.

- **Outpatient ED visits.** is an ED visit that does not result in an admission to the hospital. We can only measure outpatient ED visits in the HCCI data and therefore limit our analysis to outpatient ED visits.[22] We measure an outpatient ED visit by whether there is a claim on the outpatient file with a HCPC code corresponding to an ED visit; we allow a maximum of one outpatient ED visit per patient-day.

- **Diagnostic Tests and Imaging Procedures.** Our definition of diagnostic tests and imaging procedures follows Song et al. (2010), and is based on BETOS codes: codes beginning with T are diagnostic tests, and codes beginning with I are imaging procedures. We examine all claims files for possible diagnostic tests and imaging procedures.

- **Surgery.** We define surgeries as the sum of **inpatient surgeries** and **outpatient surgeries.** We define an **inpatient surgery** using the inpatient claims file (excluding, in the case of the HCCI data, POS codes of 31-33 since these indicate SNF). We classify an inpatient admission as having an inpatient surgery if it is associated with a "surgical DRG."[23]. We count each

[20] We do not define an analagous "SNF admission" measure because the HCCI data are not conducive to defining distinct admissions; we observe many consecutive short stays in SNFs for patients, and it is unclear whether these are distinct admissions.

[21] See http://www.dartmouthatlas.org/downloads/methods/research_methods.pdf, page 6

[22] If an ED visit results in an admission it appears on the inpatient file and can be identifed as an admission from the ED based on a "source of admission" variable. This variable however is blank / missing in HCCI.

[23] The primary source was https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareFeeforSvcPartsAB/downloads/DRGDesc10.pdf. Information on 6 DRGs (14, 16, 17, 570, 571, 572), which is not present in the above source, was added from https://www.cms.gov/Medicare/Coding/ICD10/Downloads/ICD-10-MS-DRG-v32-Definitions-Manual-Text.zip.

unique inpatient admission with a surgical DRG as one inpatient surgery. We define an **outpatient surgery** based on the HCPCS codes in the outpatient file explicitly identified as corresponding to "outpatient surgery"; we exclude any claims classified as "emergency room" claims from this definition. We restrict to a maximum of one outpatient surgery per patient-date.

Spending per encounter To measure spending per SNF day we use the above definitions of SNF spending and SNF days. To measure spending per inpatient admission or inpatient day, we use the above definition of inpatient admissions and inpatient days above; we measure inpatient spending however only counting spending on the inpatient file (i.e. not including physician spending with POS code of 21 as we do when breaking down spending by category). To measure spending per outpatient ED visit; we count all spending on the same date as the outpatient ED visit date that is on the outpatient file or is on the physician file with a POS code of 23 ("Emergency room"). For all of these measures, we take the average across enrollee months of the ratio of spending to utilization for that enrollee-month..

Preventive care We analyze the set of preventive care measures in Finkelstein et al. (2016) that we can reasonably replicate in our data. These in turn are drawn from procedures measured in the Dartmouth Atlas and the Centers for Medicare and Medicaid (CMS). These measures are typically defined as rates of any care receipt during an observation period (an enrollee-month in the baseline analysis) for a denominator of "relevant" patients. In some cases, we have to modify the denominator due to limitations of the HCCI data (e.g. coarse age bins or the inability to do a two-year "look back" period). We highlight these modifications below, which we do in parallel for both MA and TM measures so that they are internally comparable:

- **Mammogram** is defined following the Dartmouth Atlas (see http://www.dartmouthatlas.org/data/table.asp We define the denominator as women ages 65-74; due to the coarseness of the age variable in HCCI, this is a broader "risk set" than the Dartmouth Atlas denominator of women ages 67-69.
- **Diabetes screen** ("HbA1c test"), **cholesterol test** ("blood lipids test"), and **retinal eye exam (**"retinal or dilated eye exam") are defined following the Dartmouth Atlas (see http://www.dartmouthatlas.org/data/map.aspx?ind=160). For all of them the denominator (risk set) is defined as all enrollees aged 65-74 with a diagnosis of diabetes. Due to the coarseness of the age variable in HCCI, this is a slightly different "risk set" than the Dartmouth Atlas denominator of enrollees aged 65-75 with a diagnosis of diabetes. The definition of "a diagnosis of diabetes" also differs because we have only one year of data while the Dartmouth Atlas defines a diabetes diagnosis based on encounters with specific codes identifying diabetes during the year or prior year; we are able to replicate their coding exactly, but because we can only look during our one observation year, our definition is more stringent than theirs.

Information on DRG 15 was added after manual search on-line.

- **Seasonal influenza vaccine, cardiovascular screening blood test, colorectal cancer screening, pap smears, pelvic examinations,** and **prostate cancer screening** are defined following CMS' preventive care definitions (see https://www.cms.gov/Medicare/Prevention/PrevntionGenInfo/Downloads/M
9*codesseehttps* : *//www.cms.gov/Medicare/Prevention/PrevntionGenInfo/Downloads/MPS − QuickRefere*

Appendix C: Analysis of inpatient prices

In this appendix we describe our analysis for inpatient prices in MA and TM. Our objective is to compare the price of an admission at a given hospital for a given diagnosis (DRG) in MA to what this price would have been if (counterfactually) that admission had occurred under TM. For this analysis, we consider only spending on the inpatient file, and not spending on the physician file associated with the inpatient admission. We also limit our analysis to the approximately 4,000 hospitals in our baseline MA sample that, for purposes of TM reimbursement, would have been covered by Medicare's Prospective Payment System (PPS). PPS covers virtually all standard (non-specialty) hospitals; limiting ourselves to MA admissions in these hospitals excludes about 5 percent of inpatient admissions, and about 7 percent of payments to inpatient hospitals. For these standard hospitals, pricing in TM (and to the best of our understanding in MA), is based primarily on the hospital at which the admission occurs and the DRG for which the patient was admitted.

We conduct two analyses, an analysis of average price differences by state, and an analysis of average price differences by DRG (for common DRGs). They are conceptually the same, just created at different units of aggregation.

State-level prices. To arrive at a state-level average price (in either MA or TM), we calculate the average price in the state for each MA admissions in a given DRG, and then take a weighted average of prices for each DRG in the state; we use as weights the DRG's (national) share of admissions in MA.[24] As a result, any differences in average prices across states reflects price differences for a common "DRG basket."

Measuring the MA price for each MA admission is straightforward: we simply calculate total payments to hospitals for that admission, as measured in the inpatient file. Measuring the (counterfactual) TM price for each MA admission proceeds in two steps. First, we calculate the TM formula price for each MA admission. Under TM, these admissions would be reimbursed by Medicare's PPS; the PPS reimbursement formula is the product of a hospital-specific "base payment" rate times a diagnosis-specific (DRG) weight; both are publicly available from CMS.[25] We can therefore calculate, for each admission in MA, a TM formula price as a function of the hospital

[24] For a few small states, there are a number of common (national) DRGs which, in that state, have no admissions. To address this, we impute the national average price for that DRG in that missing state-DRG pair, corrected by a state-specific correction factor. The state-specific correction factor is given by the ratio of the state price and average national price for the DRGs we do observe in that state.

[25] The DRG weights can be found here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Acute-Inpatient-Files-for-Download-Items/CMS1247873.html (see file

and DRG for that admission. We compute the average, TM formula price for each DRG in the state, and then construct the state average TM formula price by taking a weighted average of prices across DRGs, using each DRG's (national) share of admissions (in MA) in that DRG as weights.

The actual, transacted TM price will not always correspond exactly to the formula TM price. For example, in certain costly cases, hospitals receive additional "outlier payments" covering 80 percent of costs beyond a threshold. In addition, if the individual is transferred to another hospital, the actual reimbursement will be below the reimbursement formula. Since in MA we observe transacted prices, in the second step, we adjust the TM formula prices to account for average differences between TM actual and TM formula prices. We calculate this adjustment factor using CMS data in which we can observe actual TM prices (i.e. payments, as we do in MA data) and can also construct TM formula prices. We calculate a state-specific adjustment factor that is the ratio of actual TM prices to formula TM prices in that state.[26] We multiply the state's average TM formula price by this state-specific adjustment factor to arrive at our estimate of the state-specific average TM price. Appendix Table 3 shows the state-specific average MA and TM prices.

DRG-level prices. The DRG-level analysis proceeds in a similar manner except that we now compute the average price for each DRG by taking a weighted-average of prices for each state in the DRG, using as weights the state's share of admissions (across all DRGs) in MA. As a result, any differences in average prices across DRGs reflects price differences for a common "state basket," which mimics the geographic distribution of MA admission across states.

The measurement of the average TM price for each DRG proceeds in the same two steps. First, we calculate each DRG's average TM formula price using the same TM formula prices for each admission that we used in the state-level analysis, but now average these across states for each DRG, using the state's share of admission as weights. Second, we adjust the TM formula price by a DRG-specific adjustment factor reflecting the DRG-specific ratio of actual TM prices to formula TM prices.[27] Appendix Table 2 shows the DRG-specific average MA and TM prices for the 20 most common DRGs.

FY_2010_FR_Table_5). The hospital base payment rates can be found in the Medicare Impact File (available here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Historical-Impact-Files-for-FY-1994-through-Present.html). The base payment rates for the hospital include hospital-specific adjustments for wage index reclassifications, indirect medical education payments, and disproportionate share payments. The HCCI data has encrypted hospital identifiers that can not be directly mapped to the publicly available data on hospital base payment rates. We are extremely grateful to Zack Cooper for providing us with a file containing these base payment rates linked to the encrypted hospital identifiers.

[26] Once again, for both actual and formula TM prices, we compute the average of admission prices by state-DRG, and then a weighted average by state, in which the weight associated to each DRG is the national share of MA admissions with that DRG.

[27] For both actual and formula TM prices, we compute the average admission prices for each state-DRG, and then a weighted average by DRG, in which the weight associated with each state is the state's share of MA admissions.

Appendix Figure 1: Mortality-Spending Relationship in TM and MA

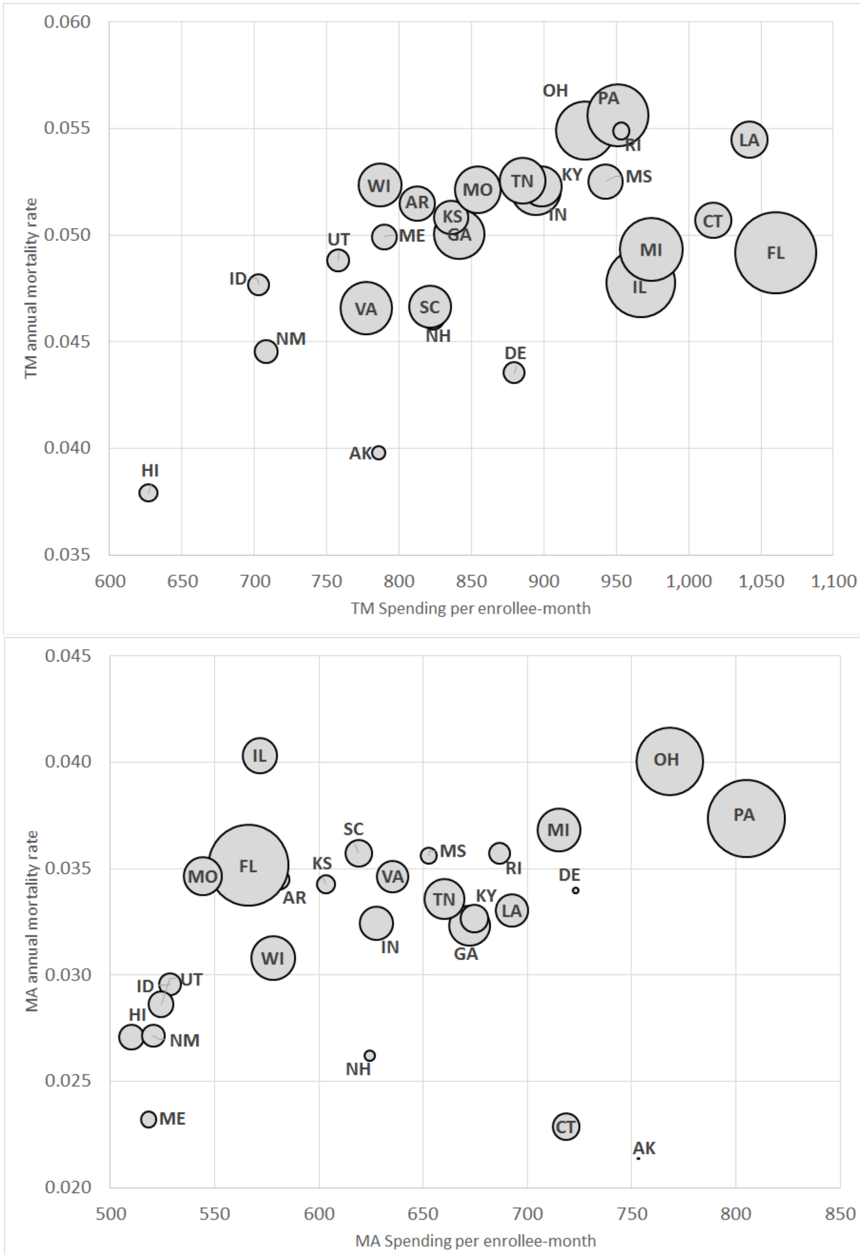

Figure shows relationship between annual mortality rate and spending for each state, separately for TM (top panel) and MA (bottom panel). In the top panel, the size of each bubble is proportional to the number of TM enrollees in the state. In the bottom panel, the size of each bubble is proportional to the number of MA enrollees in the state.

Appendix Figure 2: Propensity score distributions

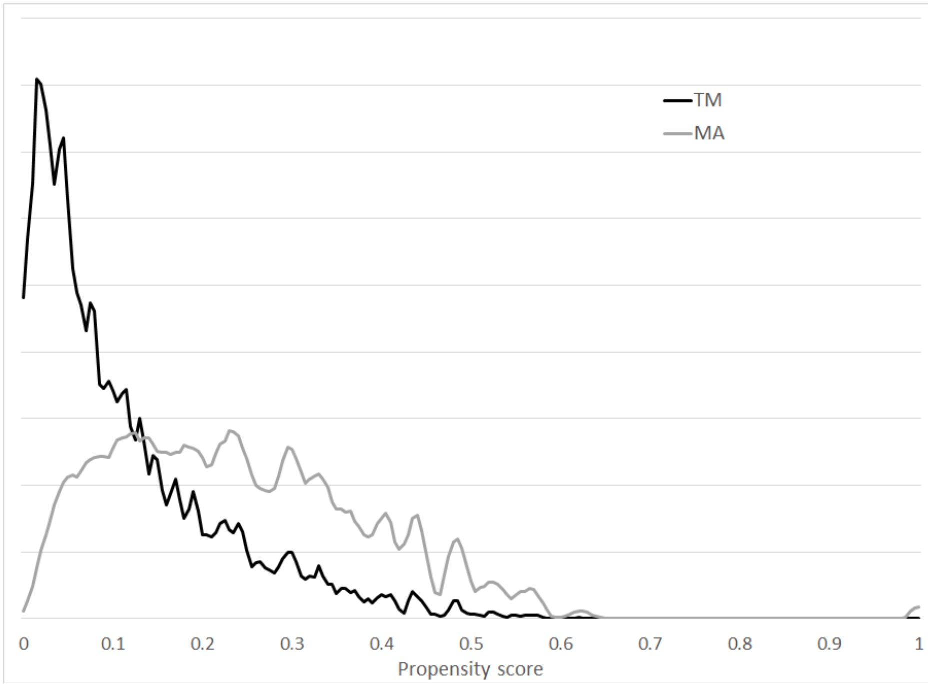

Figure shows the distribution of propensity scores in the baseline sample for the TM (black) and MA (gray) population. The figures uses row 3 of Table 9, where propensity scores are generated the predicted probability from a logit regressions of an MA indicator on dummy variables for each risk score bin (of 0.1), which is estimated county by county.

Appendix Table 1: Construction of baseline sample

	MA share (%)	HCCI insurers share of MA (%)	All HCCI	cleaned HCCI	All CMS	cleaned CMS	% Difference ((4)-(6))/(6)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All	27.7	37.2	32,422,570	31,598,737	43,738,311	39,561,941	-20.1
AL	22.2	28.0	420,588	411,913	442,324	367,197	12.2
AK	**0.8**	**58.6**	**2,593**	**2,485**	**2,804**	**2,464**	**0.9**
AZ	38.4	53.8	646,563	636,491	1,736,449	1,364,479	-53.4
AR	**14.4**	**43.2**	**271,505**	**251,796**	**283,165**	**274,613**	**-8.3**
CA	42.3	22.3	641,417	634,254	4,041,414	3,797,938	-83.3
CO	38.0	48.0	353,918	341,145	1,028,437	915,951	-62.8
CT	**21.4**	**21.0**	**210,138**	**208,868**	**238,213**	**191,344**	**9.2**
DE	**3.8**	**72.2**	**39,177**	**38,944**	**37,891**	**35,433**	**9.9**
DC	13.5	12.4	6,317	6,317	10,359	5,112	23.6
FL	**32.3**	**56.2**	**5,081,149**	**5,068,292**	**5,865,846**	**5,004,105**	**1.3**
GA	**24.2**	**69.3**	**1,737,163**	**1,713,318**	**1,818,298**	**1,644,688**	**4.2**
HI	**48.9**	**23.1**	**151,012**	**149,388**	**221,176**	**150,521**	**-0.8**
ID	**32.7**	**37.8**	**244,222**	**231,164**	**265,690**	**250,429**	**-7.7**
IL	**10.9**	**59.5**	**1,085,320**	**1,050,225**	**1,108,730**	**1,054,392**	**-0.4**
IN	**18.1**	**47.3**	**782,431**	**768,443**	**796,606**	**768,967**	**-0.1**
IA	14.6	61.4	444,220	394,347	453,505	445,774	-11.5
KS	**12.2**	**61.0**	**300,052**	**287,904**	**305,172**	**297,615**	**-3.3**
KY	**18.1**	**61.0**	**683,318**	**652,607**	**698,634**	**688,182**	**-5.2**
LA	**27.6**	**56.8**	**929,803**	**914,649**	**934,736**	**891,935**	**2.5**
ME	**15.7**	**25.0**	**87,396**	**81,183**	**91,285**	**88,281**	**-8.0**
MD	9.1	27.7	180,849	178,168	177,520	149,347	19.3
MA	25.2	8.2	140,095	138,130	194,205	102,843	34.3
MI	**18.4**	**16.1**	**445,115**	**433,695**	**453,140**	**431,439**	**0.5**
MN	49.4	9.1	320,917	298,589	334,612	333,612	-10.5
MS	**10.0**	**49.1**	**196,440**	**192,334**	**205,267**	**196,786**	**-2.3**
MO	**24.2**	**46.5**	**997,750**	**954,686**	**1,014,240**	**968,999**	**-1.5**
MT	19.7	54.4	165,679	146,583	173,066	168,162	-12.8
NE	12.9	68.0	229,845	204,094	235,535	232,357	-12.2
NV	35.3	87.8	336,761	334,850	999,167	874,172	-61.7
NH	**8.5**	**35.2**	**56,201**	**52,832**	**57,996**	**55,464**	**-4.7**
NJ	14.2	60.2	706,502	702,676	1,058,065	1,015,150	-30.8
NM	**29.7**	**20.6**	**141,871**	**125,581**	**168,847**	**127,364**	**-1.4**
NY	35.7	14.5	538,053	520,595	1,404,184	1,254,075	-58.5
NC	19.2	53.5	1,320,990	1,309,894	1,383,964	1,069,147	22.5
ND	8.9	59.7	54,486	39,421	56,494	53,139	-25.8
OH	**39.2**	**57.3**	**3,844,872**	**3,784,208**	**3,962,057**	**3,835,280**	**-1.3**
OK	16.3	47.1	182,964	172,268	417,959	391,756	-56.0
OR	47.1	14.9	205,317	198,336	402,259	372,172	-46.7
PA	**42.8**	**18.4**	**1,538,047**	**1,492,633**	**1,664,209**	**1,534,421**	**-2.7**
RI	**43.9**	**44.1**	**302,551**	**300,325**	**305,080**	**278,504**	**7.8**
SC	**16.4**	**36.1**	**395,513**	**390,084**	**406,181**	**386,203**	**1.0**
SD	8.9	68.9	78,318	66,606	82,038	81,057	-17.8
TN	**26.6**	**47.9**	**1,084,250**	**1,073,415**	**1,180,367**	**1,037,856**	**3.4**
TX	21.3	52.5	1,473,641	1,445,579	3,057,557	2,599,812	-44.4
UT	**38.7**	**51.4**	**480,094**	**463,668**	**517,464**	**486,026**	**-4.6**
VT	5.2	57.6	29,213	23,997	30,122	28,767	-16.6
VA	**15.2**	**50.4**	**760,935**	**729,157**	**787,156**	**758,468**	**-3.9**

All data except from columns (3) and (4) are from CMS. Columns (1) and (2) show the MA share of total Medicare enrollment and the HCCI insurers' share of MA enrollment, respectively. Columns (3) through (6) show counts of enrollee-months in different data sets. Columns (3) and (5) are based on the full sample of data (see columns (7) and (3) of Table 1, respectively). The "cleaned" sample in columns (4) and (6) excludes enrollees in SNP plans and masked zip codes. States that are in bold are those that are included in our baseline sample (using our criteria of counts being within 10%), and correspond to columns (8) and (6) of Table 1, respectively.

Appendix Table 2: MA-TM prices differences for most common DRGs

DRG Code (1)	DRG Description (2)	MA Admissions (3)	MA price (4)	TM price (5)	(MA-TM)/TM (6)
All DRGs (weighted by MA admission shares)		437,714	10,085	9,927	1.6%
470	Major Joint Replacement Or Reattachment Of Lower Extremity W/O Mcc	21,077	12,440	11,958	4.0%
392	Esophagitis, Gastroent & Misc Digest Disorders W/O Mcc	9,899	4,227	4,312	-2.0%
871	Septicemia Or Severe Sepsis W/O Mv 96+ Hours W Mcc	9,153	11,268	11,532	-2.3%
291	Heart Failure & Shock W Mcc	8,708	8,956	9,024	-0.7%
292	Heart Failure & Shock W Cc	8,222	5,976	6,073	-1.6%
312	Syncope & Collapse	7,467	4,452	4,478	-0.6%
690	Kidney & Urinary Tract Infections W/O Mcc	7,177	4,695	4,730	-0.7%
194	Simple Pneumonia & Pleurisy W Cc	6,606	6,031	6,039	-0.1%
310	Cardiac Arrhythmia & Conduction Disorders W/O Cc/Mcc	6,545	3,478	3,497	-0.6%
313	Chest Pain	6,175	3,352	3,379	-0.8%
247	Perc Cardiovasc Proc W Drug-Eluting Stent W/O Mcc	5,985	11,771	11,482	2.5%
190	Chronic Obstructive Pulmonary Disease W Mcc	5,919	7,277	7,260	0.2%
378	G.I. Hemorrhage W Cc	5,691	6,011	6,096	-1.4%
287	Circulatory Disorders Except Ami, W Card Cath W/O Mcc	5,655	6,358	6,376	-0.3%
641	Nutritional & Misc Metabolic Disorders W/O Mcc	5,511	4,194	4,250	-1.3%
193	Simple Pneumonia & Pleurisy W Mcc	5,076	8,706	8,693	0.1%
683	Renal Failure W Cc	4,869	6,487	6,415	1.1%
192	Chronic Obstructive Pulmonary Disease W/O Cc/Mcc	4,847	4,354	4,377	-0.5%
191	Chronic Obstructive Pulmonary Disease W Cc	4,771	5,855	5,847	0.1%
293	Heart Failure & Shock W/O Cc/Mcc	4,621	4,254	4,285	-0.7%

Table reports average prices for a hospital admission in TM and MA for the top 20 DRGs, and overall across all DRGs(not limited to the top 20). Averages are computed for each DRG using a common (MA) "basket" of state admission shares. Sample is a subset of our baseline sample; it is limited to all MA inpatient admissions to hospitals that are paid (by CMS) under prospective payment system (PPS).

Appendix Table 3: MA-TM price differences, by state

State (1)	MA Admissions (2)	MA price (3)	TM price (4)	(MA-TM)/TM (5)
AR	4,338	9,420	9,011	4.5%
CT	2,886	11,497	12,771	-10.0%
DE	772	10,794	11,470	-5.9%
FL	103,324	10,291	9,830	4.7%
GA	27,154	10,299	9,892	4.1%
HI	1,339	13,279	13,176	0.8%
ID	2,079	10,305	9,757	5.6%
IL	18,480	10,182	10,333	-1.5%
IN	11,649	9,700	9,542	1.6%
KS	5,079	9,421	9,490	-0.7%
KY	12,883	9,677	9,688	-0.1%
LA	20,502	9,947	9,834	1.2%
ME	932	10,340	10,736	-3.7%
MI	8,485	10,096	10,943	-7.7%
MO	16,814	9,631	9,541	0.9%
MS	3,865	9,716	9,458	2.7%
NH	538	10,856	10,692	1.5%
NM	1,560	10,863	11,187	-2.9%
OH	87,631	9,562	9,871	-3.1%
PA	34,338	11,169	10,403	7.4%
RI	5,116	11,571	12,067	-4.1%
SC	6,663	10,007	10,283	-2.7%
TN	23,664	9,755	8,909	9.5%
UT	5,813	9,589	9,319	2.9%
VA	13,167	9,787	9,885	-1.0%
WI	18,609	10,345	10,488	-1.4%

Table reports average prices for a hospital admission in TM and MA for each state in our baseline sample (except Alaska which is omitted because it had too few inpatient admissions for us to report). Averages are computed for each state using a common (MA) "basket" of DRG admission shares. Sample is a subset of our baseline sample; it is limited to all MA inpatient admissions to hospitals that are paid (by CMS) under prospective payment system (PPS)..