Are All Managed Care Plans Created Equal? Evidence from Random Plan Assignment in Medicaid*

Michael Geruso† Timothy J. Layton‡ Jacob Wallace§

August 25, 2020

Abstract

Exploiting random assignment of Medicaid beneficiaries to managed care plans, we identify plan-specific effects on healthcare utilization. Auto-assignment to the lowest-spending plan generates 30% lower spending than if the same enrollee were assigned to the highest-spending plan, despite *identical* cost-sharing. Effects via quantities, rather than differences in negotiated prices, explain these patterns. Rather than reducing "wasteful" spending, low-spending plans cause broad reductions in the use of medical services—including low-cost, high-value care—and worsen beneficiary satisfaction and health. Supply side tools circumvent the classic trade-off between financial risk protection and moral hazard, but give rise instead to a cost/quality trade-off.

*We thank Marika Cabral, Michael Chernew, David Cutler, Josh Gottlieb, Ben Handel, Jon Kolstad, Neale Mahoney, Tom McGuire, Mark Shepard, Ben Sommers, Amanda Starc, and Bob Town, as well as seminar participants at AHEC, ASHEcon 2016, the BU/Harvard/MIT Health Economics Seminar, the Chicago Booth Junior Health Economics Summit, Harvard Medical School, Hunter College, Rice University/Baker Institute, and Stanford for useful feedback and suggestions. We also thank the New York State Department of Health (and particularly Greg Allen, Jason Ganns, Chang Byun, Foster Gesten, Hyun Jae Kang, and Pat Roohan) for assistance in providing and interpreting the data. Layton and Wallace gratefully acknowledge funding support from the National Institute of Mental Health (Grant No. T32-019733) and the National Science Foundation Graduate Research Fellowship (Grant No. DGE 1144152), respectively, as well as the Laura and John Arnold Foundation and the Agency for Healthcare Research and Quality (Grant No. K01-HS25786-01). Geruso gratefully acknowledges support by grant P2CHD042849, Population Research Center, awarded to the Population Research Center at The University of Texas at Austin by the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The conclusions and opinions presented in here are those of the authors and do not necessarily reflect those of the New York State Department of Health or any funder.

†UT-Austin and NBER. Email: mike.geruso@austin.utexas.edu

‡Harvard University and NBER. Email: layton@hcp.med.harvard.edu

§Yale School of Public Health. Email: jacob.wallace@yale.edu

1 Introduction

Regulated competition between private health plans is becoming the dominant form of social health insurance in the United States (Gruber, 2017). In 2017, 54 million Medicaid beneficiaries (69%) and 19 million Medicare beneficiaries (33%) were enrolled in a private managed care plan (Kaiser Family Foundation, 2017, 2019). In the same year, almost $500 billion of the $1.3 trillion spent on public health insurance programs went to private managed care plans.

An important feature of publicly-subsidized health insurance markets is that beneficiaries typically have a choice among competing private plans. The availability of consumer choice makes it difficult to establish important facts about the plans with which governments are contracting. For example, do plans with observably lower levels of spending (or higher quality) relative to their competitors in the same markets *causally* reduce spending (or increase quality)? Or are performance differences due to the fact that plans attract different mixes of enrollees? This selection bias hampers efforts to understand what role plans can play in constraining healthcare spending growth and to characterize the trade-offs that accompany the cost-saving approaches of plans. The identification challenge echoes that in other contexts inside and outside of healthcare—e.g., estimating physician effects (Doyle, Ewer and Wagner, 2010); hospital effects (Doyle et al., 2015; Hull, 2020); place effects (Finkelstein, Gentzkow and Williams, 2016, 2019); teacher effects (Chetty, Friedman and Rockoff, 2014*a*,*b*); school effects (Aizer, Currie and Moretti, 2007); and neighborhood effects (Chetty, Hendren and Katz, 2016; Chetty and Hendren, 2018*a*,*b*).

Understanding the extent to which competing plans can impact healthcare spending, consumer satisfaction, and enrollee health is of first-order importance in the management of these publicly-subsidized private markets. In some cases, observational measures of plan spending per beneficiary lead to statutory transfers to or from plans (e.g., penalties tied to Minimum Loss Ratio requirements or subsidies provided via risk-sharing arrangements). In others, observed plan quality is publicly-reported to consumers, factored into plan incentive payments (e.g., plan benchmark payments linked to star ratings in Medicare Advantage), or incorporated into policies that encourage enrollment in "higher quality" plans (e.g., default assignment policies in Medicaid). And yet, the prior literature has not established the extent to which plans—rather than the enrollees they attract—can, in fact, influence these metrics.

In this paper, we identify the causal effects of the health plan in which a beneficiary enrolls on her healthcare utilization, the quality of care received, and proxies for satisfaction and health. The context of our analysis is Medicaid Managed Care (MMC), the privatized system through which most Medicaid beneficiaries receive benefits today. In particular, we study the second-largest MMC market in the United States, New York City (NYC), where ten plans competed for enrollees during our study period. Like many state Medicaid programs, beneficiaries in New York who do not actively choose a plan within the designated choice period are *randomly assigned* to one. Using administrative data that cover nearly 70,000 randomly-assigned enrollees, we estimate *causal* plan differences in healthcare spending and outcomes for the ten plans operating in NYC.

In our setting, all plans are required to provide care at zero marginal cost to beneficiaries. It is therefore an ideal context for studying whether various *non-cost-sharing* plan features (e.g., networks, negotiated provider rates, patient follow-up and medication adherence programs, etc.) can constrain healthcare spending. In contrast, nearly all of the prior econometric literature studying how health plans affect utilization and health outcomes has focused on consumer cost-sharing provisions like copays, coinsurance, and deductibles. But a modern health plan is more than a set of consumer-facing prices, and our analysis sheds new light on the range of impacts generated by supply-side (non-cost-sharing) plan features. To facilitate a transparent comparison between our results and results from cost-sharing studies including the RAND Health Insurance Experiment (Manning et al., 1987) and more recent quasi-experimental work (Brot-Goldberg et al., 2017), we focus our analysis on the types of outcomes that have been the focus of this prior literature. These include overall service utilization and spending, utilization of high- and low-value care, conventional measures of healthcare quality, and surrogate health outcomes like avoidable hospitalizations.

As our first main result, we document statistically and economically significant *causal* variation in spending across plans. If an individual enrolls in the lowest-spending plan in the market she will generate about 30% less in healthcare spending than if the same individual enrolled in the highest-spending plan in the market. We show that risk-adjusted observational measures and causal estimates of plan spending effects are correlated, but find that the risk-adjusted measures tend to overstate causal differences in spending across plans. Plans that attract healthier patients thus do more to constrain spending—i.e., provide less care—consistent with a classic adverse selection model, where sicker individuals select plans providing more care. This fact has important implications for the use of observational measures of spending and quality as a basis for regulatory rewards or penalties.

After establishing important differences between risk-adjusted (OLS) plan spending effects and causal (IV) estimates, we investigate which factors drive the bottom-line causal differences. First, we find that almost all services are marginal. That is, lower spending plans tend to provide less of nearly everything. This includes inpatient and outpatient visits, primary care physician office visits, and high-value/cost effective drugs. Second, unlike in other markets (Cutler, McClellan and Newhouse, 2000; Gruber and McKnight, 2016; Cooper et al., 2019), differences in provider prices do not explain the differences in healthcare spending across plans in our setting. In a decomposition, prices account for very little of the cross-plan spending differences.

Instead, spending differs because enrollees in low-spending plans use less care, with much of the utilization gap driven by the extensive margin. Importantly (and similar to the effects of deductibles in Brot-Goldberg et al., 2017), utilization reductions do not seem to focus on "low-value" care or "waste": We estimate that low-spending plans reduce utilization of high-value drugs used to treat diabetes, asthma, and severe mental illnesses, as well as high-value screenings for diabetes, cancer, and sexually transmitted infections.

Finally, we show that the low-spending plans also increase avoidable hospitalizations and decrease consumer satisfaction, as measured by the propensity of auto-assigned enrollees to switch out of their plan post-assignment. These results suggest a clear trade-off between spending and beneficiary satisfaction and health.

This paper contributes to a nascent literature attempting to estimate health plan effects in settings where health plans differ in more than cost-sharing parameters. This complements contemporaneous research on Medicare Advantage by Abaluck et al. (2020), Medicaid Managed Care in South Carolina by Garthwaite and Notowidigdo (2019), and health plans serving the non-elderly, non-Medicaid population by Handel et al. (2018). Our work also contributes to the literature on optimal insurance design in the presence of moral hazard. We provide new evidence on how an under-studied set of health plan features (those not related to cost-sharing) constrain spending, adding to a smaller recent literature concerned with these features (see, e.g., Curto et al. (2017); Layton et al. (2019); Wallace (2019)). Consistent with Garthwaite and Notowidigdo (2019), we show that there is substantial *causal* heterogeneity across plans in spending and utilization that arises without any differences in consumer cost-sharing exposure.

Our findings complement a large literature extending back to the RAND health insurance experiment (Manning et al., 1987) that documents how consumer prices impact healthcare utilization. In RAND, and the studies that have followed, patient cost-sharing has proven to be a blunt instrument, affecting the use of low- and high-value services alike (Brot-Goldberg et al., 2017). These findings sparked interest in whether managed care tools offer a scalpel that can target inefficient spending and better manage the high-cost patients responsible for the majority of spending. But our results, along with prior work studying managed care in Medicare (Curto et al., 2017), indicate that supply-side tools exhibit many of the same features and limitations as demand-side tools. Their impacts on healthcare spending are blunt. They indiscriminately reduce utilization, limiting both high- and low-value care rather than targeting "waste." In another similarity to the effects of consumer cost-sharing (as found in Brot-Goldberg et al., 2017), lower-spending managed care plans in our setting do not appear to generate savings by steering patients to lower-cost providers or lowering negotiated prices.

Lastly, our work highlights how supply side tools can achieve spending reductions while circumventing the classic trade-off between financial risk protection and moral hazard noted by Zeckhauser (1970) and Pauly (1974). The spread of plan effects we estimate are similar to the utilization difference between the 0% and 95% coinsurance rate treatment arms in the RAND HIE. Thus, significantly constraining healthcare spending need not require exposing consumers to out of pocket spending. But there is no "free lunch" here, as we also document that these spending reductions come at the cost of beneficiary satisfaction and, ultimately, health.

2 Data and Setting

2.1 Medicaid Managed Care in New York

New York State is similar to the broader US in its reliance on private managed care organizations (MCOs) to deliver Medicaid benefits to the majority of its Medicaid beneficiaries.[1] New York is typical in that Medicaid beneficiaries may choose plans from a range of carriers that include national for-profits, local for-profits, and local non-profits, though we are not permitted to identify specific plans in our analysis. We focus on New York City, which is comprised of five counties where enrollment in managed care is mandatory, and which contains about two-thirds of the state's Medicaid

[1] See Appendix A for additional detail.

population. Restricting attention to a single large city allows us to identify differences across managed care plans operating in the same healthcare market.

2.2 Auto-assignment to Plans

For our study period (2008-2012), beneficiaries in New York City had 30, 60, or 90 days to actively choose an MCO. In excess of 90 percent of beneficiaries did so. Our study design focuses on the beneficiaries who did not choose within the required timeframe and were automatically assigned to a plan, a policy known as "auto-assignment." These auto-assigned enrollees were randomly allocated across eligible plans with equal probability. We observe the assignment and use it as an instrument for enrollment in a given plan, as we discuss below in Section 3.

After notification of auto-assignment, each beneficiary had three months to switch plans without cause before a nine-month lock-in period began. This is the primary explanation for imperfect compliance (i.e., a first stage effect of assignment on enrollment smaller than 1.0). Additional institutional details regarding auto-assignment are available in Appendix A and are documented in Wallace (2019), which examines the effect of Medicaid managed care provider networks in New York.

We construct our "auto-assignee sample" with the following restrictions. First, we restrict the sample to beneficiaries aged 18 to 64. We exclude individuals aged 65 and older because they are excluded from managed care. We remove beneficiaries below age 18 because children are often non-randomly auto-assigned to their parents' plans. Second, we exclude Medicaid beneficiaries with family members in a Medicaid managed care plan at the time of auto assignment and beneficiaries who were enrolled in a managed care plan in the year prior to assignment. Plan assignments for these beneficiaries are automatic, but not random. Third, we restrict to beneficiaries with at least six months of post-assignment enrollment in Medicaid to allow us to observe plan effects on spending, utilization, and quality outcomes.

In primary analyses we restrict attention to the initial six months post-assignment. Enrollment is high and stable until six months and then drops off precipitously (see Appendix Figure A1). This is due to high levels of churn in the Medicaid program combined with a NY regulation guaranteeing Medicaid eligibility for six months following the beginning of an MMC enrollment spell. We show robustness of our main results to expanding the sample to include additional months in Appendix D. The expanded-sample results are nearly identical.

These sample restrictions leave us with 65,596 beneficiaries in five boroughs and ten plans. The final "auto-assignee" sample includes 285 county $\times$ year $\times$ month (the unit of randomization) cohorts of observations. In some instances, we make comparisons to the full NYC MMC population from the sample period, which includes the 1,011,169 18- to 64-year-old "active choosers" who were not randomized to a plan. Tables 1 and A1 describe these samples. Internal validity does not require that auto-assignees and active choosers are similar. Nonetheless, one might be concerned that auto-assignees are healthier and less-engaged with the healthcare system, so that estimates reflect impacts on consumers who use little care. Table A1 shows that auto-assignees do differ somewhat from active-choosers, being more likely to be black males. But on overall healthcare spending, the groups appear similar. In fact, auto-assignees use slightly *more* care than active-choosers. The IV analysis thus estimates plan effects on patterns of healthcare utilization, health, and satisfaction among individuals actively using care.[2]

2.3 Administrative data and outcomes

We obtained detailed administrative data from the New York State Department of Health (NYSDOH) for the non-elderly New York Medicaid population from 2008 to 2012. The data contain beneficiary-level demographic and enrollment data linked to healthcare claims for services covered by fee-for-service Medicaid (FFS) and private MCOs. The MCO enrollment data include an indicator for whether a beneficiary made an active plan choice or was auto-assigned, and, for auto-assignees, the plan of assignment. Monthly plan enrollment data allow us to observe whether beneficiaries remained in their assigned plans. We construct beneficiary-month level outcomes related to healthcare use and spending.

Healthcare use, prices, and spending. Claims data include information on providers, transaction prices, procedures, and quantities. When measuring healthcare use and spending we include all services paid for by the managed care plans and by fee-for-service Medicaid, which carves out certain services from managed care financial responsibility. Also, most beneficiaries spend a few months enrolled in the FFS program prior to choosing or being assigned to a managed care plan, allowing us to observe utilization under a common fee-for-service regime prior to randomization. This enables powerful balance tests.

[2]Anywhere where we compare results for active-choosers and auto-assignees, we re-weight the active-choosers to match the auto-assignees on demographics and detailed, pre-enrollment healthcare utilization.

Healthcare quality. We measure healthcare quality by adapting access measures developed by the Secretary of Health and Human Services (HHS) for the adult Medicaid population. We determined whether beneficiaries complied with recommended preventive care, measured as the frequency of flu vaccination for adults ages 18 to 64, breast cancer screenings, cervical cancer screenings, and chlamydia screenings in women. We also examined the frequency of avoidable hospitalizations (a surrogate health outcome), operationalized as admission rates for four conditions: diabetes short-term complications, chronic obstructive pulmonary disease (COPD) or asthma in adults, heart failure, and asthma in younger adults. We use additional measures of potentially high- and low-value care that follow recent contributions in the literature (Schwartz et al., 2014; Brot-Goldberg et al., 2017).

Willingness-to-Stay. Because Medicaid enrollees do not pay a premium (price) for enrolling with any of the plans in the market, we cannot measure beneficiary willingness-to-pay for one plan versus another. Instead, we assume beneficiaries' preferences are revealed through their subsequent plan choices (voting with their feet). While switching rates are low, enrollees are not locked-in to their assigned plans: For the first three months after assignment they may switch for any reason, after which they can switch for "good cause." As we discuss in Section 4.3, we measure willingness-to-*stay* as the likelihood that a randomly-assigned enrollee remains in her assigned plan.

3 Empirical Framework and First-Stage

3.1 Econometric Model

Our main empirical goal in this paper is to measure the causal effect of enrollment in health plan $j \in J$ on outcomes at the beneficiary (i) level. We follow Finkelstein, Gentzkow and Williams (2016) in modeling a data generating process for healthcare spending in which log spending (Y_{ij}) is determined by a plan component (γ_j), a person-level fixed effect (ξ_i), time-varying observables (X_{it}), and a mean zero shock (ϵ_{ijct}) that may be influenced by both beneficiary and plan:[3]

$$Y_{ijct} = \nu X_{it} + \gamma_j + \xi_i + \epsilon_{ijct}. \tag{1}$$

To recover plan effects, γ_j, we estimate Equation 1 as a regression at the individual-level, com-

[3]FGW decompose spending into beneficiary and place effects, holding plan (fee-for-service Medicare) fixed. We (effectively) decompose spending into beneficiary and plan effects, holding place fixed.

bining the ξ_i and ϵ_{ijct} terms into a compound error term μ_{ijct}:

$$Y_{ijct} = \rho + \psi_{ct} + \nu X_{ict} + \sum_{j=1}^{9} \gamma_j \mathbf{1}[\text{Plan\_j}_{\ ict}] + \mu_{ijct}. \tag{2}$$

In these regressions, an observation is a beneficiary-month. The regressors of interest are indicators for enrollment in month t in each of the ten plans competing in the New York City market (with one plan as the omitted category). Fixed effects ψ_{ct} for month $t \times$ county c of enrollment are included in all specifications. The X vector of individual controls is described below.

To address the endogeneity of beneficiaries sorting themselves across plans—correlation between plan choice and μ_{ijct}—we exploit random assignment. We restrict to individuals who were randomly auto-assigned to plans and instrument for plan *enrollment* indicators with plan *assignment* indicators. There are ten plans that receive auto-assigned enrollees during our time period, requiring nine first-stage regressions (with plan 10 omitted):

$$\begin{aligned} \text{Plan\_1}_{\ ict} &= \alpha_1 + \phi_{1ct} + \delta_1 X_{ict} + \sum_{j=1}^{9} \lambda_{1j} \mathbf{1}[\text{Assigned\_j}_{\ ict}] + \eta_{1,ict} \\ &\vdots \\ \text{Plan\_9}_{\ ict} &= \alpha_9 + \phi_{9ct} + \delta_9 X_{ict} + \sum_{j=1}^{9} \lambda_{9j} \mathbf{1}[\text{Assigned\_j}_{\ ict}] + \eta_{9,ict}. \end{aligned} \tag{3}$$

We use the nine first-stage regressions to predict enrollment in each plan. For each auto-assigned enrollee, only one of the plan assignment variables will be equal to one. The coefficient λ_{kj} captures the probability that an individual auto-assigned to plan j will be enrolled in plan k during any given month, relative to the omitted plan. For each first-stage regression, a λ_{kj} equal to one for $k = j$ and equal to zero for $k \neq j$ would indicate perfect compliance. The second stage estimating equation uses the vector of predicted enrollment values ($\widehat{\mathbf{Plan}_{ijct}}$) from the first-stage regressions:

$$Y_{ijct} = \rho + \psi_{ct} + \nu X_{ict} + \sum_{j=1}^{9} \gamma_j \mathbf{1}[\widehat{\text{Plan\_j}}_{\ ict}] + \mu_{ijct}. \tag{4}$$

This IV strategy results in estimates of the plan effects, γ_j, that use only variation in enrollment due to quasi-random auto-assignment.

3.2 First-Stage and Instrument Validity

Figure 1B plots λ_{jj} for each plan—roughly, the probability that a beneficiary who is auto-assigned to a plan is enrolled in that plan after assignment. In the six-months post-assignment, beneficiaries spend more than ninety percent of the time in their assigned plan. The overall first-stage F-statistic is reported in Table 1 and exceeds 7,000. Table A2 lists all of the first-stage coefficient estimates, λ_{kj}. The high rate of compliance implies the local-average treatment effects recovered by IV are unlikely to differ much from average treatment effects for the auto-assignee sample.

Figure 1A presents a series of randomization tests to assess the independence assumption to the extent possible, using information on pre-determined characteristics like demographics, as well as pre-randomization medical expenditure. To test for correlations between assignment and pre-determined characteristics, each baseline characteristic is regressed on nine indicators for beneficiaries' assigned plans (omitting one plan to prevent perfect collinearity). We perform this regression separately for auto-assignees and a random sample of active-choosers of equal size to the auto-assignee sample to equalize statistical power across the two groups.

In the left panel of Figure 1A, we plot plan effect coefficients from the active-chooser regressions as hollow circles and coefficients from the auto-assignee regressions as solid circles. In the right panel, we plot for each dependent variable the p-value from an F-test that the plan effects are jointly different from zero, again separately for the active-chooser and auto-assignee samples.[4] High p-values are consistent with random assignment. The results in the figure provide strong evidence of balance across plans for the auto-assignees, with plan effects tightly clustered around zero for all pre-determined characteristics. p-values exceed 0.05 for all but one characteristic. In contrast, the analogous estimates for the active-choosers show that plan "effects" on pre-determined characteristics are large, and each characteristic is predicted by plan choice with $p < 0.05$. The imbalance among active-choosers indicates that the *lack* of statistical imbalance among the auto-assignees is not due to noisy or uninformative observables. It also suggests that selection would be an important confounder in the absence of quasi-random assignment. Below in Figure 4 and Table A4, we provide evidence that there is no differential attrition as a function of plan of assignment, another important fact to establish for our identification.

[4]Tabular versions of these results are in Table A3.

4 Results

4.1 Healthcare spending

We start by presenting results for each plan's causal effect on spending relative to an omitted plan, using the IV regression in Equation 4. Panel (a) of Figure 2 reports the main result—plan effects on monthly $\log(\text{spending} + 1)$ from the IV regression. The plotted coefficients reveal substantial heterogeneity in spending and utilization across plans. We estimate that the highest-spending plan, Plan D, spends 11.5% more than the omitted plan. The lowest-spending plan, Plan I, spends 20.3% less than the omitted plan. This implies a range in spending of 31.8%, with six plans (A, B, C, G, H, I) spending significantly less than the omitted plan (X), two plans (E and F) exhibiting spending levels similar to the omitted plan, and one plan (D) having significantly higher spending. The same panel also reports coefficients from a regression in which the dependent variable is an indicator for any utilization in the month. This regression reveals similar patterns, with lower-spending plans exhibiting lower probabilities of non-zero utilization each month.

These patterns are robust to alternative specifications and constructions of the dependent variable. Table 1 reports results with and without controls for pre-determined characteristics. We estimate similar variation in plan effects when the outcome is parameterized as the inverse hyperbolic sine of spending or Windsorized spending levels (Table A5), and when we aggregate spending over the entire six-month enrollment spell, rather than analyzing monthly outcomes (Table A6). In Appendix C, we show that these results are not merely reflecting a temporary disruption of care at the time of assignment; month-by-month event studies (Figure A2) show differences that persist for the entire enrollment spell.

The range of these estimates is large. For example, the range of our plan effects corresponds to 2.5 times the size of the spending difference between plans featuring care at zero marginal cost versus a high deductible (Brot-Goldberg et al., 2017).[5] Yet, our estimates are considerably smaller than the

[5] A collage of additional evidence indicates that these plan effects are not merely the artifacts of differential reporting across plans. E.g., the Office of the Inspector General examined New York Medicaid managed care plans in 2012 and found a large range (about a 30 percentage point span) in medical loss ratios across plans (OIG, 2015). As we discuss in detail in Appendix B, the underlying cost data for the OIG report differs from the claims data we use and so provides independent corroboration. Carve-outs provide further corroboration: A minority of services for MMC enrollees are carved-out and paid by the state via FFS. When we examine each spending component (FFS and MMC) separately, we show that the patterns of FFS claims—in which the plans themselves have no reporting role—track the patterns of MMC claims. Thus, reduced utilization in low-spending plans is observed even where data originate with the state rather than with any plan. See Appendix B for full detail and additional supporting evidence.

observational, cross-sectional differences in plan spending. To better understand this relationship, Panel (b) of Figure 2 plots plan effects identified via random assignment in the IV sample against plan effects (estimated via OLS) that compare the spending of enrollees making active plan choices. Both regressions include rich controls (risk adjusters) for observable enrollee characteristics, including deciles of *ex-ante* spending from the period prior to the beneficiary entering MMC, during which all beneficiaries were enrolled in FFS. Further, the active-chooser sample is reweighted to match the distribution of observables in the auto-assignee IV sample to provide the most consistent comparison the data allow. These coefficients are also reported in Tables 1 and A7.

Figure 2 indicates a noisy relationship between the observational and causal estimates. On average, enrolling in a plan with high risk-adjusted spending among active choosers (x-axis) will cause an enrollee to have higher spending (y-axis). But this average relationship masks substantial heterogeneity: The size and even ranking of plan effects varies in the two sets of estimates, indicative of substantial selection across plans. On average, the observed selection is adverse: Higher-spending enrollees opt into plans with larger positive causal effects on spending. Such selection suggests that conventional cross-sectional comparisons of spending or other outcomes across plans would be difficult to interpret, as differences will be driven by both causal plan effects *and* residual selection. This is true even when adjusting for a rich set of observables that include prior healthcare spending.

4.2 Mechanisms

To better understand mechanisms behind the spending differences, we divide the ten plans into three sets based on the IV spending effects: low (Plans A, B, C, G, H, I), medium (Plans E, F, and X), and high (Plan D). The grouping aids with statistical power, as well as with tractability of the comparisons. In the modified IV regression, the endogenous variables are indicators for enrollment in any plan in each set, and the instruments are indicators for assignment to any plan in each set.[6] Because Plan D (the single outlier high-spending plan) is so different from the others in terms of overall spending, we focus on the low versus medium coefficients in the main text. We report our consistent, but noisier, estimates for the high plan in Appendix Figures A3 and A4.

We focus on two classes of potential mechanisms that mirror the mechanisms most often explored in the literature examining the effects of consumer cost-sharing parameters like deductibles and coin-

[6]That is, $Y_{ict} = \rho + \psi_{ct} + \nu X_{ict} + \gamma_{\text{Low}} \mathbf{1}[\widehat{\text{Low Plan}}_{ict}] + \gamma_{\text{High}} \mathbf{1}[\widehat{\text{High Plan}}_{ict}] + \mu_{ict}$.

surance rates. First, we explore the role of negotiated provider prices, which could operate by lower-spending plans steering consumers to different, lower-priced providers or by lower-spending plans negotiating lower prices with the same providers. Second, we explore whether the tools used by the lower-spending plans to reduce utilization are "blunt" in that they tend to reduce all types of care or "sharp," in the sense of targeting and cutting low-value services. In Appendix D, we also briefly study the role of provider network breadth, finding little relationship between network size and plan spending effects in this setting.

Prices The similarity of plan effects on total spending and plan effects on an indicator for any utilization (Figure 2) suggests that quantity differences may be more important than negotiated price differences in explaining spending differences in this context. Figure 3 investigates the role of prices in greater detail. In it we plot the median log prices for the medium- versus low-spending plans for thousands of services. A price here is the paid amount at the level of the DRG for inpatient admissions (Panel a) and at the level of the procedure code for outpatient services (Panel b). The figure shows that prices appear very similar across medium- and lower-spending plans. Systematically higher prices in the medium-spending group would appear as a vertical shift of the cloud of points above the 45 degree line. No shift is evident. Analogous figures for high- versus medium-spending plans are shown in Figure A3, revealing a similar pattern.

To decompose exactly how much of the spending differences can be accounted for by prices, we next re-price all claims as if all plans transacted at a common set of prices. We then re-calculate enrollee spending using the re-priced claims and re-run the IV analysis.[7] The results are displayed in Table A5. For ease of comparison, Panel (c) of Figure 3 plots our main IV plan effects against plan effects estimated on the price-standardized data. Re-pricing has almost no effect on our estimates of plan spending coefficients, indicating that price differences cannot account for the spending differences we observe.

In the spending panel of Figure 4, we show that denied claims also do not appear to be important for explaining plan spending effect differences. Each row of the figure corresponds to a separate regression. In the first row of the first panel, we report the main result at the group level: Low-spending plans, on average, generate 16% less spending than medium-spending plans. To illustrate the role of denials, we reprice each denied (zero paid) claim as if it had been paid at a common-across-

[7] See Appendix C for full detail on the repricing, which follows Cooper et al. (2019).

plans price, and then re-estimate the effect of low- versus medium-spending plans. This change does not alter the difference in spending we calculate between the two groups of plans.

Blunt versus sharp In the RAND HIE and the quasi-experimental studies that have followed it, patient cost-sharing has proven to be a blunt instrument, with deductibles and coinsurance affecting use of low- and high-value services alike. Are the reductions in spending generated by managed care similarly blunt or are these better targeted?

Figure 4 investigates. The figure shows that reductions in low-spending plans occur across all services: inpatient admissions, pharmacy, outpatient care, office visits, lab services, and dental care. The most-rationed services were office visits and hospital outpatient services.

So far, our findings do not rule out the possibility that low-spending plans invest in high-value treatments that make people healthier and decrease the need for costly inpatient and outpatient hospital treatments. To investigate this, we examine two sets of potentially high-value services that could produce spending offsets: drugs and preventive services.

Figure 4 show no evidence that low-spending plans invest more in high-value drugs or preventive services. With respect to drugs, we focus on a set of maintenance drugs used to treat chronic conditions. Specifically, we estimate plan effects on diabetes drugs, statins, anti-depressants, anti-psychotics, anti-hypertensives, anti-stroke drugs, asthma drugs, and contraceptives. Rather than increase utilization, low-spending plans decrease utilization of most of these drugs, though some reductions are statistically insignificant. This is inconsistent with the idea that lower-spending plans use scalpel-like tools to reduce inefficient spending while improving or maintaining provision of high-value care: For many of these drugs non-adherence can result in health deterioration and expensive hospitalizations.

We also study the effects of enrolling in a low-spending plan on the use of other, non-drug, high-value services. Figure 4 analyzes four measures of compliance with recommended care developed by Health and Human Services for Medicaid enrollees: the prevalence of HbA1c testing, breast cancer screening rates, cervical cancer screening rates, and chlamydia screening rates. For each of the measures except breast cancer screening, we find that enrollment in a low-spending spending plan significantly reduces the use of recommended preventive care.

We next investigate whether low-spending plans target low-value care for cutting. We estimate the effects of enrolling in a low-spending plan on the use of a variety of potentially low-value services, including inappropriate abdominal imaging, back imaging, and colorectal screening (Schwartz et al., 2014; Charlesworth et al., 2016). We find no evidence that low-spending plans reduce the use of these low-value services, with suggestive evidence that rates of low-value care may actually be higher.

In sum, there is no indication that low-spending plans achieve savings by promoting high-value care and achieving offsets or by targeting low-value care for elimination. Instead, similar to what happens when consumers face a high deductible, supply-side managed care tools appear to constrain virtually all types of care.

4.3 Trade-offs

Finally, to the extent possible in our data, we evaluate the effects of enrolling in low-spending plans on welfare-relevant outcomes beyond utilization. The literature on demand-side cost-sharing has shown that cost-sharing involves a trade-off between risk protection and moral hazard (Zeckhauser, 1970). Researchers and policymakers have also been interested in the effects of demand-side cost-sharing on health. The evidence on this relationship has been mixed.

In the Medicaid setting, beneficiaries enrolling in lower-spending plans are not subject to cost-sharing. So the financial risk trade-off is absent. There may, however, still be a trade-off between satisfaction/utility and plan spending. We study this trade-off by estimating differences in the probability that an individual assigned to a low- versus medium-spending plan opts to stay in that plan post-assignment rather than switch to a different plan. We call our measure of this probability "willingness-to-stay," and assume that differential switching behavior across randomly-assigned plans is correlated with enrollees' experienced utility across plans.

Figure 4 shows that people are less likely to stay in lower-spending plans. Appendix Figure A4 shows that beneficiaries are likewise *more* likely to stay in the high-spending plan versus the medium-spending plans. To give a finer view of these results, in Appendix Figure A5, we plot plan-level estimates of willingness-to-stay against the plan effects on spending—first for all beneficiaries and then stratified according to health status. The relationship is clear, with higher-spending plans having higher estimates of willingness-to-stay. It is also clear that sicker beneficiaries—those who use more care and so have more experience with their plans—drive this relationship. These results provide evidence of a real trade-off between plan spending and beneficiary satisfaction.

Finally, to the extent possible in our data, we examine whether there are measurable health effects of the tools used by low-spending plans to constrain cost. We do this with a standard surrogate health outcome that can be constructed from claims data: hospitalizations that are potentially avoidable given appropriate treatment and management of a set of common conditions. The measures were developed by the Agency for Healthcare Research and Quality (AHRQ) for the Medicaid population. (See Appendix B for details.) Figure 4 shows that enrollees in the low-spending plans are 16% *more* likely to have an avoidable hospitalization despite having lower utilization for most other types of care. This result suggests that the tools used by low-spending plans to constrain costs could have negative consequences for beneficiary health.

5 Conclusion

Our results are important for understanding the potential for managed care to constrain healthcare spending growth. We show that the baskets of rationing devices implicit in managed care can have spending and utilization impacts significantly larger than what could be accomplished by exposing consumers to high deductibles and reasonable coinsurance and copays.

Importantly, rationing via managed care reduces spending without exposing consumers to financial risk, circumventing the classic trade-off between financial risk protection and moral hazard noted by Zeckhauser (1970) and Pauly (1974). These findings are particularly relevant for public insurance programs—including the low-income segments of HIX Marketplaces and Medicare—where policymakers have been reluctant to expose low-income consumers to financial risk.

However, these spending reductions appear to come with a utility cost. Willingness to remain enrolled in a plan is negatively related to that plan's cost savings. And cost reductions are blunt—reducing utilization of all types of care, lowering traditional measures of healthcare quality, and increasing the likelihood of adverse health events.

References

Abaluck, Jason, Mauricio M. Caceres Bravo, Peter Hull, and Amanda Starc. 2020. "Mortality Effects and Choice Across Private Health Insurance Plans." National Bureau of Economic Research Working Paper 27578.

Aizer, Anna, Janet Currie, and Enrico Moretti. 2007. "Does Managed Care Hurt Health? Evidence from Medicaid Mothers." *Review of Economics and Statistics*, 89(3): 385–399.

Brot-Goldberg, Zarek C, Amitabh Chandra, Benjamin R Handel, and Jonathan T Kolstad. 2017. "What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics." *The Quarterly Journal of Economics*, 132(3): 1261–1318.

Charlesworth, Christina J, Thomas HA Meath, Aaron L Schwartz, and K John McConnell. 2016. "Comparison of low-value care in Medicaid vs commercially insured populations." *JAMA internal medicine*, 176(7): 998–1004.

Chetty, Raj, and Nathaniel Hendren. 2018*a*. "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects." *The Quarterly Journal of Economics*, 133(3): 1107–1162.

Chetty, Raj, and Nathaniel Hendren. 2018*b*. "The impacts of neighborhoods on intergenerational mobility II: County-level estimates." *The Quarterly Journal of Economics*, 133(3): 1163–1228.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014*a*. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates." *American Economic Review*, 104(9): 2593–2632.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014*b*. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American economic review*, 104(9): 2633–79.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz. 2016. "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment." *American Economic Review*, 106(4): 855–902.

Cooper, Zack, Stuart V Craig, Martin Gaynor, and John Van Reenan. 2019. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured." *Quarterly Journal of Economics*, 134(1): 51–107.

Curto, Vilsa, Liran Einav, Amy Finkelstein, Jonathan Levin, and Jay Bhattacharya. 2017. "Healthcare Spending and Utilization in Public and Private Medicare." National Bureau of Economic Research Working Paper 23090.

Cutler, David M, Mark McClellan, and Joseph P Newhouse. 2000. "How does managed care do it?" *The Rand journal of economics*, 526–548.

Doyle, Joseph J, John A Graves, Jonathan Gruber, and Samuel A Kleiner. 2015. "Measuring returns to hospital care: Evidence from ambulance referral patterns." *Journal of Political Economy*, 123(1): 170–214.

Doyle, Joseph J, Steven M Ewer, and Todd H Wagner. 2010. "Returns to physician human capital: Evidence from patients randomized to physician teams." *Journal of health economics*, 29(6): 866–882.

Ericson, Keith, and Amanda Starc. 2015. "Measuring Consumer Valuation of Limited Provider Networks." *American Economic Review Papers and Proceedings*, 105(5): 115–119.

Finkelstein, Amy, Matthew Gentzkow, and Heidi L Williams. 2019. "Place-based drivers of mortality: Evidence from migration." National Bureau of Economic Research.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics*, 131(4): 1681–1726.

Garthwaite, Craig, and Matthew Notowidigdo. 2019. "Plan Value-Added: Evaluating Medicaid Managed Care Plans Using Random Assignment." Working Paper.

Gruber, Jonathan. 2017. "Delivering Public Health Insurance through Private Plan Choice in the United States." *Journal of Economic Perspectives*, 31(4): 3–22.

Gruber, Jonathan, and Robin McKnight. 2016. "Controlling health care costs through limited network insurance plans: Evidence from Massachusetts state employees." *American Economic Journal: Economic Policy*, 8(2): 219–50.

Handel, Benjamin, Jonathan Holmes, Jonathan Kolstad, and Kurt Lavetti. 2018. "Insurer Innovation and Health Care Efficiency: Evidence from Utah." Working paper.

Hull, Peter. 2020. "Estimating Hospital Quality with Quasi-experimental Data."

Kaiser Family Foundation. 2017. "Medicare Advantage 2017 Spotlight: Enrollment Market Update." http://files.kff.org/attachment/Issue-Brief-Medicare-Advantage-2017-Spotlight-Enrollment-Market-Update.

Kaiser Family Foundation. 2019. "10 Things to Know About Medicaid Managed Care." https://www.kff.org/medicaid/issue-brief/10-things-to-know-about-medicaid-managed-care/.

Layton, Timothy, Nicole Maestas, Daniel Prinz, and Boris Vabson. 2019. "Healthcare Rationing in Public Insurance Programs: Evidence from Medicaid." National Bureau of Economic Research Working Paper 26042.

Lewin Group. 2012. "Evaluating Encounter Data Completeness." https://www.ccwdata.org/documents/10280/19002254/evaluating-encounter-data-completeness.pdf.

Manning, Willard G, Joseph P Newhouse, Naihua Duan, Emmett B Keeler, and Arleen Leibowitz. 1987. "Health insurance and the demand for medical care: evidence from a randomized experiment." *The American economic review*, 251–277.

OIG. 2015. "The Medicaid Program could have achieved savings if New York applied medical loss ratio standards similar to those established by the Affordable Care Act." *Department of Health and Human Services Office of the Inspector General*.

Pauly, Mark V. 1974. "Overinsurance and public provision of insurance: The roles of moral hazard and adverse selection." *The Quarterly Journal of Economics*, 44–62.

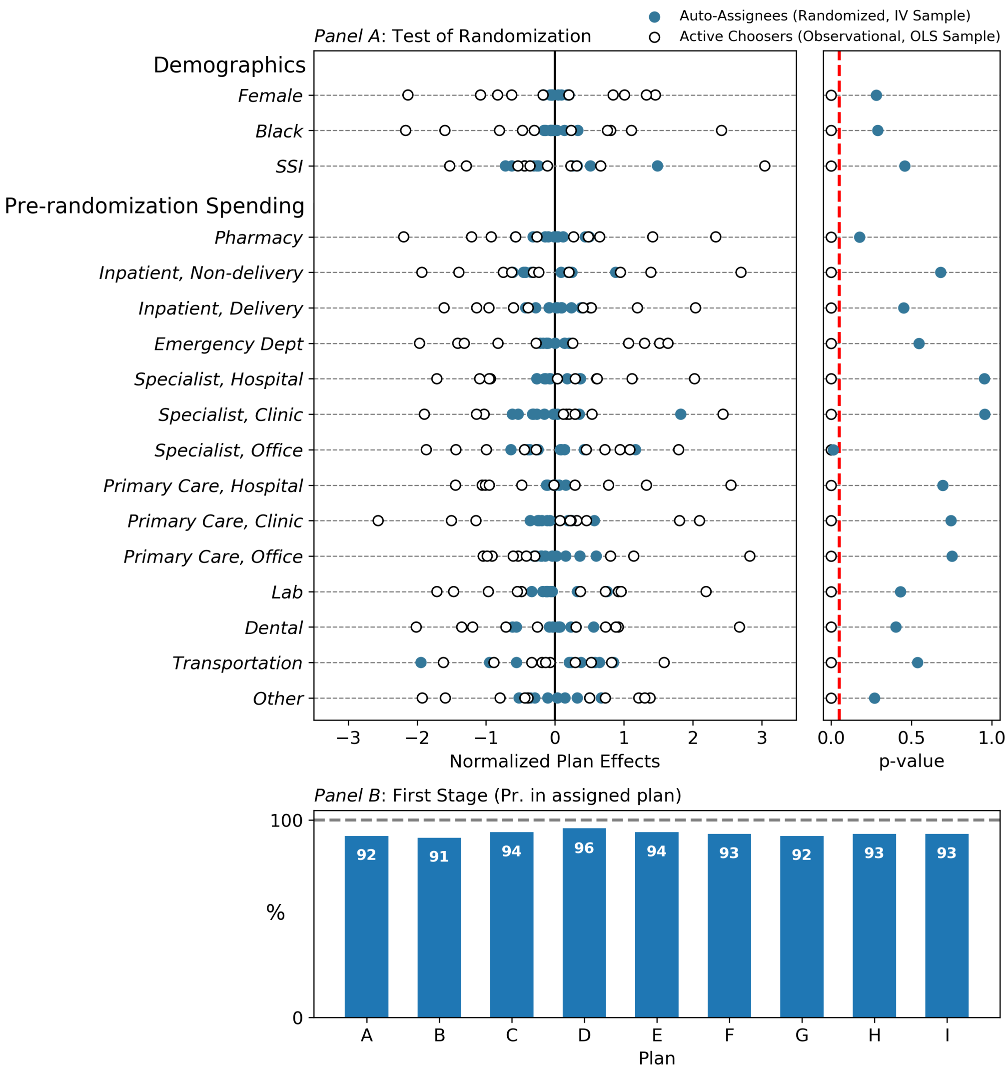
Schwartz, Aaron L, Bruce E Landon, Adam G Elshaug, Michael E Chernew, and J Michael McWilliams. 2014. "Measuring low-value care in Medicare." *JAMA internal medicine*, 174(7): 1067–1076.

Sparer, Michael. 2012. "Medicaid Managed Care: Costs, Access, and Quality of Care." Robert Wood Johnson Foundation Research Synthesis Report 23.

Wallace, Jacob. 2019. "What Does a Provider Network Do? Evidence from Random Assignment in Medicaid Managed Care." *SSRN Electronic Journal*. Available at SSRN: https://ssrn.com/abstract=3544928 or http://dx.doi.org/10.2139/ssrn.3544928.

Zeckhauser, Richard. 1970. "Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives." *Journal of Economic theory*, 2(1): 10–26.

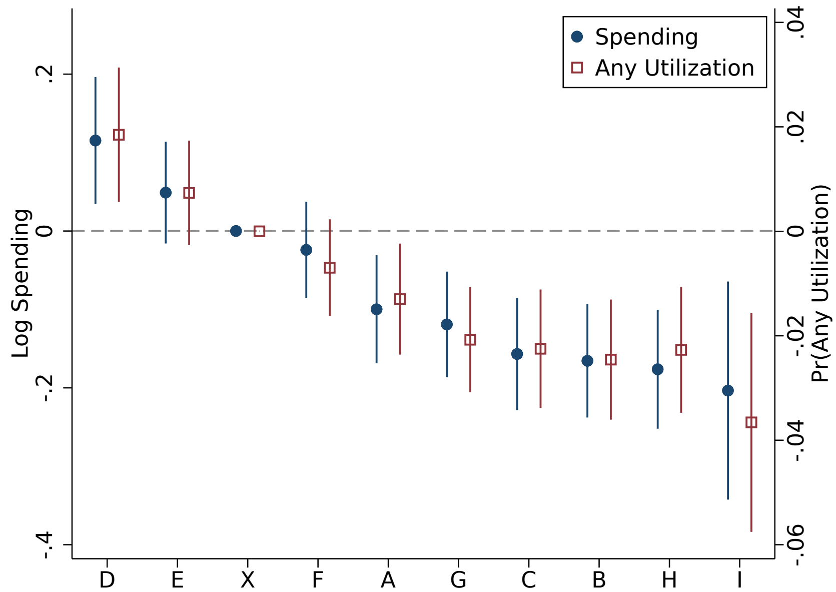
Figure 1: First Stage and Instrument Balance on Predetermined Characteristics



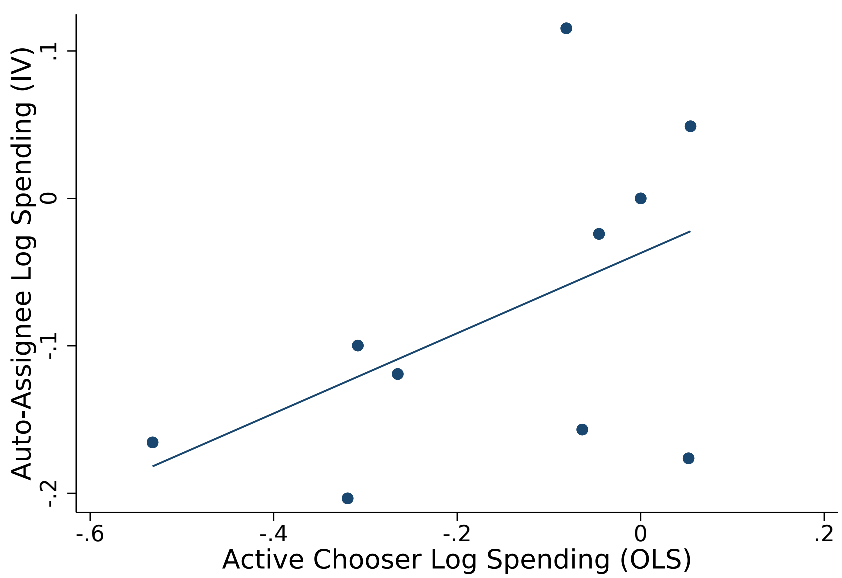
Note: Figure displays a balance test for the randomization in Panel (a) and first stage regression coefficients in Panel (b). Pre-determined characteristics include demographics and healthcare utilization in FFS Medicaid prior to randomized auto-assignment to a managed care plan. Each enrollee spent a pre-period (often a few months, once retroactive enrollment is included) enrolled in the FFS program prior to choosing or being assigned to a managed care plan. For the balance test, two samples are used: the main IV analysis sample of auto-assignees (AA) and a same-sized random subsample of active choosers (AC) for comparison. On the left side of Panel (a), each pre-determined characteristic is regressed on the set of indicators for the assigned plan (for auto-assignees) or for the chosen plan (for active choosers), and the plan effects are plotted. Separate regressions are run for the AA and AC groups, so that each horizontal line plots plan coefficients from two regressions. The plan effects are demeaned within the AA and AC groups separately, and then scaled by the same factor (the standard deviation of the combined set of demeaned plan effects). Hence, the scales (not displayed) differ for each dependent variable but are identical for the AA and AC regressions within a dependent variable. Tighter groupings of estimated plan coefficients indicate smaller differences across plans in the characteristics of enrollees. In the right side of Panel (a), we show the p-values from F-tests that the plan effects in these regressions are jointly different from zero. Tabular versions of these results are in Table A3. Large p-values are consistent with random assignment. Small p-values indicate selection on observables. The vertical dashed line is at p=0.05. In the bottom panel, bar heights correspond to coefficients from the first stage regressions (Eq. 3), in which observations are enrollee-months, the coefficient plotted is on an indicator for assignment to plan j, and the dependent variable is enrollment in plan j. Bar heights can be interpreted as approximately the fraction of months auto-assignees remain in their plan of assignment. Table A2 reports all first stage coefficients.

Figure 2: Main Results: IV Plan Effects on Healthcare Utilization

(a) Causal spending and use effects (IV)



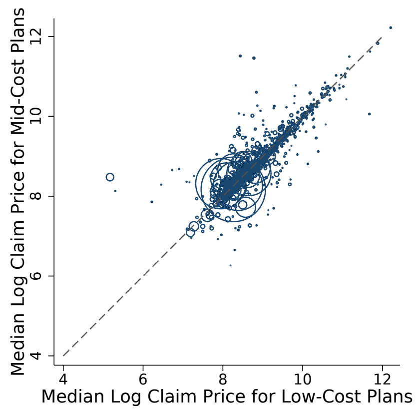
(b) Observational vs. causal spending effects



Note: Figure displays the main results of the paper—plan effects on healthcare utilization identified by random plan assignment. Panel (a) plots IV coefficients corresponding to Eq. 4, where the dependent variable is log(total healthcare spending +1) on the left axis or an indicator for any spending in the enrollee-month on the right axis. Plan of enrollment is instrumented with plan of assignment. Coefficients are relative to the omitted plan, X. For the plot, plans are ordered by their spending effects. Whiskers indicate 95% confidence intervals. Standard errors are clustered at the county $\times$ year $\times$ month-of-assignment level. This is the level at which the randomization operates. Panel (b) compares the same IV estimates from panel (a) with the observational differences in spending across plans within the active chooser sample. Active chooser (observational) differences are estimated as OLS coefficients in a regression of log total monthly spending on a full set of plan indicators, as in Eq. 2. The active chooser sample is reweighted to match the IV sample on observables, including FFS healthcare utilization prior to managed care enrollment. Person-level controls are identical in the OLS and IV specifications. See the notes to Tables 1 and A7 for tabular forms of these results and for complete details on the control variables and reweighting.

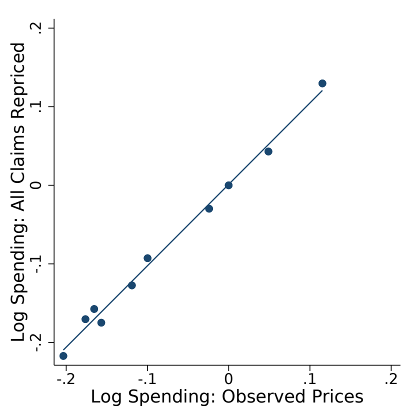
Figure 3: Transaction Price Differences Do Not Account for Spending Differences

(a) Inpatient Price Comparison (DRGs) for Medium- vs. Low-Spending Plans



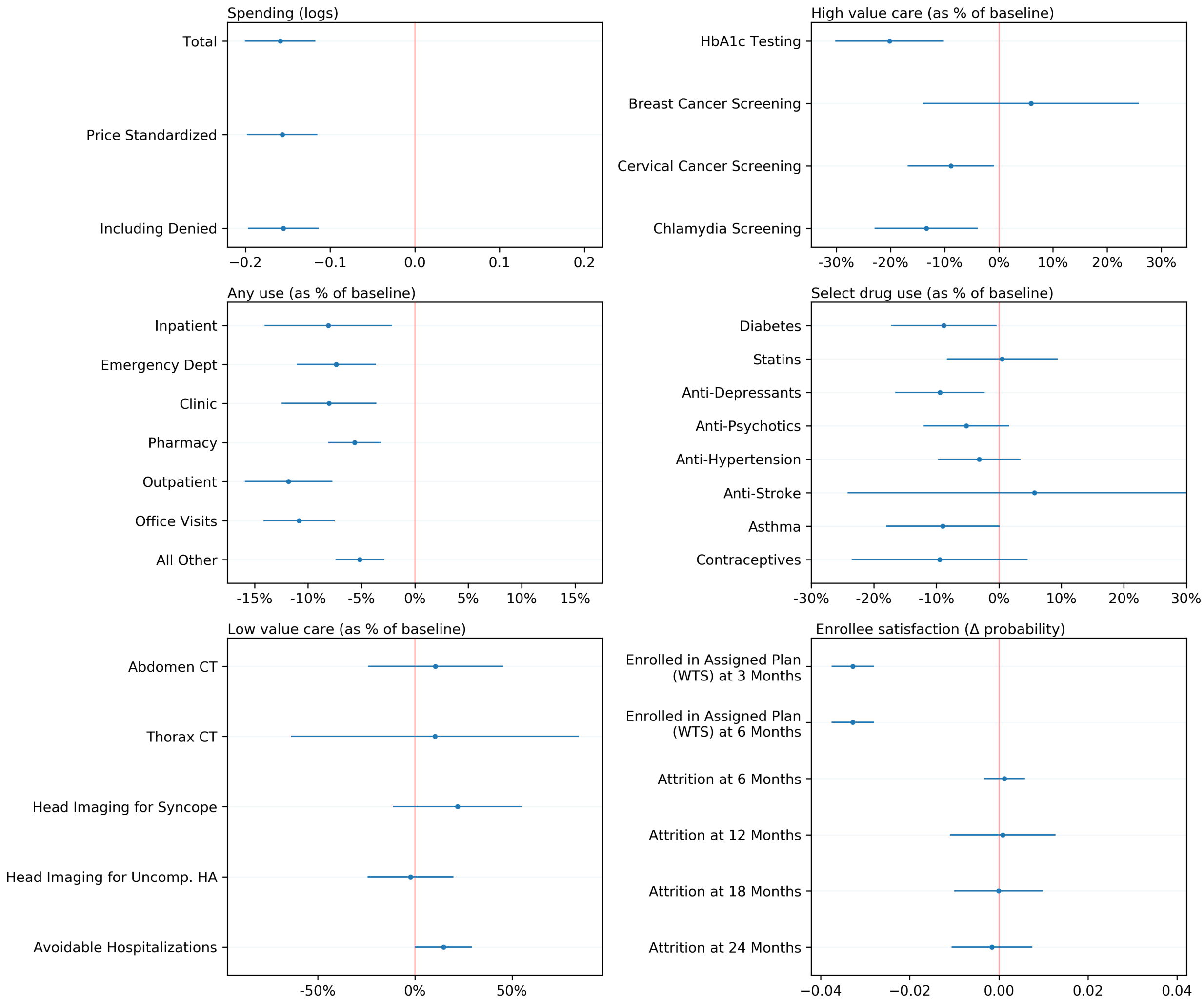
(b) Outpatient Price Comparison (HCPCS) for Medium- vs. Low-Spending Plans

(c) Normalized Spending: Repricing All Claims to Common Price List



Note: Figure shows the minor role played by transaction prices in explaining spending differences across plans. The top two panels divide plans into high-, medium-, and low-spending groups as described in the text. We focus on medium- and low-spending plan groups as the high spender is a single plan outlier. Figure A3 shows analogous comparisons for high- versus medium- and high- versus low-cost plans. Panel (a) plots the log of median prices for all inpatient admissions with common support in our data among medium- and low-spending plans. Each circle in Panel (a) is a diagnosis-related group (DRG), and marker size is proportional to frequency in our claims data. Panel (b) plots the analogous price comparison for outpatient claims, using the Healthcare Common Procedural Coding System (HCPCS). Panel (c) reverts to a plan-level analysis and reprices all claims to a common set of prices across all plans and then re-estimates the main IV specification for plan effects on log spending. The plan spending effects for the repriced data are plotted along the vertical axis, against the main (non-repriced) IV estimates along the horizontal axis.

Figure 4: Mechanisms



Note: Figure shows spending and utilization in low-spending plans compared to medium-spending plans across various categories and service settings. Plans are divided into three sets: low- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high- (Plan D) spending. We estimate a modified version of the IV regression in Eq. 4 in which the endogenous variables are indicators for enrollment in any plan in each set: $Y_{ict} = \rho + \psi_{ct} + \nu X_{ict} + \gamma_{\text{Low}} \mathbf{1}[\widehat{\text{Low Plan}_{ict}}] + \gamma_{\text{High}} \mathbf{1}[\widehat{\text{High Plan}_{ict}}] + \mu_{ict}$. Medium spending is the omitted category. The instruments are indicators for *assignment* to any plan in each set. We focus here on coefficients on the low-spending group indicator (γ_{Low}), because the high spender is a single plan outlier. (Figure A4 reports the analogous results for the single high-spending outlier.) Labels to the left within each panel describe the dependent variable. Coefficients are plotted with 95% confidence intervals. Coefficients in the first panel are effects on log spending. In the next four panels, coefficients are divided by the mean of the dependent variable in the omitted group to allow placing multiple outcomes on the same scale. In the last panel, which describes willingness-to-stay enrolled in the assigned plan (WTS) and attrition out of sample, the dependent variables are indicators and the coefficients are not scaled. For example, a WTS coefficient of -0.03 would correspond to an effect in which enrollment in a low-spending plan—in place of a medium-spending plan—increased the probability of switching plans by three percentage points. For a complete tabulation of all regression results displayed in the Figure, see Tables A8, A9, A10, and A11.

Table 1: Main Results: Plan Effects on Spending and Plan Switching

	Summary Statistics		Regression Results						
			OLS Spending		IV Spending			Willingness-to-Stay	
Plan	Number of Auto-Assignees (IV Sample) (1)	% of Active Choosers Selecting Plan (2)	Log Spending (3)	Log Spending (4)	Log Spending (5)	Log Spending (6)	Any Spending in Enrollee-Month? (7)	Enrolled in Assigned Plan at 3 mos? (8)	Enrolled in Assigned Plan at 6 mos? (9)
A	8,514	9.9	-0.264**	-0.273**	-0.101*	-0.100**	-0.013*	-0.029**	-0.041**
			(0.024)	(0.015)	(0.042)	(0.035)	(0.005)	-(0.004)	-(0.004)
B	7,816	6.7	-0.564**	-0.449**	-0.176**	-0.166**	-0.025**	-0.047**	-0.067**
			(0.037)	(0.024)	(0.042)	(0.037)	(0.006)	-(0.004)	-(0.004)
C	6,206	6.3	0.046*	-0.145**	-0.166**	-0.157**	-0.022**	-0.009*	-0.015**
			(0.019)	(0.014)	(0.045)	(0.036)	(0.006)	-(0.004)	-(0.005)
D	2,628	18.5	0.216**	-0.059**	0.171**	0.115**	0.018**	0.014**	0.019**
			(0.013)	(0.009)	(0.050)	(0.041)	(0.007)	-(0.004)	-(0.005)
E	6,753	11.6	0.101**	-0.055**	0.058	0.049	0.007	-0.025**	-0.028**
			(0.019)	(0.013)	(0.040)	(0.033)	(0.005)	-(0.003)	-(0.004)
F	8,074	18.1	0.290**	-0.101**	-0.011	-0.024	-0.007	-0.001	-0.006
			(0.019)	(0.012)	(0.036)	(0.031)	(0.005)	-(0.003)	-(0.004)
G	8,449	5.7	0.027	-0.204**	-0.134**	-0.119**	-0.021**	-0.041**	-0.056**
			(0.023)	(0.014)	(0.041)	(0.034)	(0.005)	-(0.004)	-(0.004)
H	7,087	6.8	-0.056*	-0.079**	-0.156**	-0.176**	-0.023**	-0.020**	-0.030**
			(0.026)	(0.020)	(0.046)	(0.038)	(0.006)	-(0.004)	-(0.005)
I	1,384	3.5	-0.499**	-0.354**	-0.164+	-0.203**	-0.037**	-0.030**	-0.046**
			(0.022)	(0.017)	(0.084)	(0.071)	(0.011)	-(0.006)	-(0.008)
X	8,685	12.9							
County x Year x Month FEs			X	X	X	X	X	X	X
Person-Level Controls				X		X	X	X	X
First Stage F-Statistic					7,143	7,043	7,043		
Obs: Enrollees								65,596	65,596
Obs: Enrollee X Months			6,067,014	6,067,014	393,576	393,576	393,576		

Note: Table displays summary statistics and main results. Column 1 reports counts of auto-assignees. When aggregated over the study period, plans received different numbers of auto-assignees depending on whether the plans were offered in the county and eligible for auto-enrollees at the time of assignment (see Appendix A). Column 2 reports the percent of active choosers selecting each plan. Remaining columns report OLS or IV regression results, where dependent variables are indicated in the column headers. In columns 3–7, plan regressors correspond to the plan of current enrollment in the enrollee-month. For the IV regressions (columns 5–7), these are instrumented with plan of initial assignment. Kleibergen-Paap F statistics from the first stage are reported. See Table A2 for first stage coefficients. In columns 8 and 9, the dependent variable is an indicator for remaining in the auto-assigned plan at three and six months post-assignment, respectively. Observations are enrollee $\times$ months in columns 3 through 7 and enrollees in columns 8 and 9. OLS regressions include only active-choosers; see Table A12 for additional OLS results that pool the active chooser and auto-assignee (IV) samples. Person-level controls include: sex, 5 race categories, deciles of spending in FFS prior to MMC enrollment, and 47 age categories (single years from 18 to 64). All regressions control for county $\times$ year $\times$ month-of-assignment and the count of months since plan assignment/plan enrollment, both as saturated sets of indicators. Standard errors in parentheses are clustered at the county $\times$ year $\times$ month-of-assignment level. This is the level at which the randomization operates. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

For Online Publication

Appendix for:
Are All Managed Care Plans Created Equal: Evidence from Random Plan Assignment in Medicaid

A Medicaid Managed Care in New York

New York State began experimenting with managed care in Medicaid in 1967. In 1997, New York obtained a Section 1115 waiver from the Department of Health and Human Services that authorized a statewide Medicaid Managed Care program utilizing private carriers in place of a traditional fee-for-service program. This program was voluntary in the 1980s and expanded into a mandatory program in the 1990s and 2000s.[1] Under mandatory managed care, beneficiaries are required to join a managed care plan operated by a for-profit or not-for-profit third party organization.

A.1 Broader Nationwide Context

During the study sample period New York State was similar to the national mean in its use of private managed care organizations to administer Medicaid enrollee benefits. According to CMS, as of July 2011 (toward the end of the study sample period), about three quarters of New York State's Medicaid beneficiaries were enrolled in a managed care program. The Kaiser Family Foundation reports that as of 2014, 77 percent of the US Medicaid population was enrolled in a Medicaid Managed Care plan, with 39 states using MCOs to deliver Medicaid benefits.

A.2 Auto Assignment in NYC

There are two exceptions to the auto-assignment policies described in Section 2. First, New York takes into account family member enrollment, defaulting beneficiaries into their family member's plan. Second, beneficiaries who were enrolled in a managed care plan in the year prior to assignment are reassigned to their previous plan.[2] Beneficiaries assigned on the basis of family members or prior enrollment are flagged and removed from our analysis sample.[3]

For our study period, in New York City beneficiaries had 30, 60, or 90 days to make an active choice. In practice, the gap we observe between enrollment and auto-assignment (see Appendix Figure A1) is often in excess of 90 days. During our study period (and today), Medicaid beneficiaries were retroactively enrolled upon successful application—a mechanism intended to cover recent unpaid medical bills that would have been covered by Medicaid. From a 2011 NY Medicaid policy document: "the retroactive eligibility period ... begins on the first day of the third month prior to the month in which the individual applied for Medicaid and ends on the date the individual applies for Medicaid."[4] Thus, although auto-assignment happens within 90 days of successful application, the

[1] The shift to mandatory managed care took place via county-by-county "enrollment mandates." The mandates initially applied only to children and TANF adults, but were expanded to include disabled Medicaid beneficiaries (Sparer, 2012).

[2] Preferential assignment to a prior plan does not apply if the beneficiary's prior plan was a partial capitation plan, a low quality plan, or a plan without further capacity.

[3] Auto assignments on the basis of family members of prior enrollees are not directly separately identified in the data. We adopt a conservative approach to removing these beneficiaries, flagging and dropping anyone with a case (family) member in their file at the time they are auto-assigned. We also remove beneficiaries with any managed care enrollment in the year prior to auto-assignment.

[4] The document, which includes additional details on New York's retroactive eligibility policy, is available here: https://www.health.ny.gov/health_care/medicaid/reference/mrg/june2011/pages495.6-8.pdf. (Accessed 8/17/2020)

observed enrollment spell often extends back prior to application, including the retroactive period as well. When taking this retroactive eligibility period into account, beneficiaries could be enrolled in the fee-for-service (FFS) Medicaid program for as long as 6 months prior to auto-assignment (which we often observe, as reported in Appendix Figure A1). Beneficiaries could also be enrolled for longer than six months prior to assignment if their assignment occurs due to a new MMC enrollment mandate for their eligibility group. While MMC enrollment mandates were in effect for most populations in NYC prior to the beginning of our study period, some small groups were transitioned at some point during the period. These groups would have 30, 60, or 90 days to make an active choice *from the date the mandate kicks in*, not from the date they applied for Medicaid. Given that some of these individuals could have been enrolled in Medicaid for years prior to the implementation of an MMC enrollment mandate for their group, it is possible for these beneficiaries to have pre-assignment enrollment periods much longer than six months.

Plans qualify as eligible for assignment based on a yearly composite measure that incorporates state-specific quality measures, Consumer Assessment of Healthcare Providers and Systems (CAHPS) responses, Prevention Quality Indicators (PQIs), and regulatory compliance measures. Prevention Quality Indicators (PQIs) are a set of measures developed by the Agency for Healthcare Research and Quality to evaluate the quality of care for "ambulatory care sensitive conditions." These are conditions for which good outpatient care can prevent hospitalizations or complications. Because plans do not necessarily qualify for random assignment over our entire study period and are not always available in all counties, we treat a beneficiary's county-by-year-by-month of assignment as the unit of randomization.

A.3 Auto Assignee Sample Sizes by Plan

The sample size of auto-assignees is not identical across plans for several reasons. First, as noted above, plans qualify to receive auto-assignees based on a yearly performance composite that measures plan-level quality, consumer satisfaction, and regulatory compliance. Plans that don't qualify are ineligible to receive auto-assignees during the specified period. Second, some of the plans in our sample do not service Staten Island, one of the five boroughs of New York City, and so will not receive auto-assignees that reside there.

In addition to the two factors above, there was a merger of two of the plans in our sample. The merger, which took place in the final year of our study (2012), led to all enrollees in the acquired plan being transferred to the acquiring plan. Since this was not a voluntary plan switch, for the set of auto-assignees that were in the acquired plan, we recoded their plan of assignment to be the acquirer beginning the month of the acquisition.

B Data

B.1 Administrative data and outcomes

All managed care plans are required to submit standardized encounter data for the services they provide and the NYSDOH has linked this data to the claims they pay directly through the FFS program. Thus, our claims data include both MMC and FFS components. An evaluation by the Lewin Group indicated that the New York data was ready for use in research (Lewin Group, 2012).

We use the validated administrative data from the NYSDOH to construct a series of outcomes including enrollee spending, utilization of medical services and drugs, healthcare quality (including avoidable hospitalizations), plan satisfaction, and the likelihood of re-enrolling in Medicaid. All of these outcomes are either used by policymakers to regulate plans, publicly-reported to enrollees during the plan choice process, or both. We briefly describe the details of these outcomes below.

- **Categories of service.** We use an algorithm provided by the New York State Department of Health to classify administrative healthcare claims into mutually-exclusive categories of service. The state's algorithm takes into account the claim type, provider category of service, provider specialty code, rate code (a New York data element used to identify the broad type of service provided), procedure code (e.g., CPT, HCPCS, ICD), modifier code, and enrollee age.

- **Drug classification.** We use Truven Health Analytics Red Book to classify the pharmaceutical claims in our data. Red Book groups claims into mutually-exclusive buckets based on the National Drug Code (NDC). Our drug groups are supersets of REDBOOK therapeutic classes. Diabetes includes: Anti-diabetic agents, Sulfonylureas; Anti-diabetic agents, misc; Anti-diabetic agents, Insulins. Statins include: Anti-hyper-lipidemic Drugs. Anti-depressants include: Psychother, Anti-depressants. Anti-psychotics include: Psychother, Tranq/Antipsychotic; ASH, Benzodiazepines; Anticonvulsant, Benzodiazepine. Anti-hypertension includes: Cardiac, ACE Inhibitors; Cardiac, Beta Blockers; Cardiac, Alpha-Beta Blockers. Anti-stroke includes: Coag/ Anticoag, Anticoagulants. Asthma/COPD includes: Adrenals Comb, NEC.

- **Healthcare quality.** We construct three sets of healthcare quality measures. First, we determine whether beneficiaries comply with recommended preventive care. Second, we examine the rate of avoidable hospitalizations. And, third, we measure the prevalence of low value care.

  Preventive care. We examined whether beneficiaries complied with recommended flu vaccinations for adults ages 18 to 64, breast cancer screenings, cervical cancer screenings, and chlamydia screenings in women. These measures follow the specifications of the Medicaid Adult Core Set HEDIS measures but do not include any continuous enrollment restriction for inclusion. The Breast cancer screening measure determines the percentage of women ages 50 to 65 who had a mammogram. The cervical cancer screening measure determines the percentage of women ages 21 to 64 who were screened for cervical cancer. Chlamydia screening determines the percentage of sexually active women 18 to 24 who were tested for chlamydia. The HbA1c measure determines the percentage of diabetic adults ages 18 to 64 who had a hemoglobin A1c test.

  Avoidable hospitalizations. Avoidable hospitalizations follow the specifications of the Medicaid Adult Core Set HEDIS measures. PQI-01 counts the number of inpatient hospitalizations for diabetes short term complications for adults ages 18 to 64. PQI-05 counts the number of inpatient hospitalizations for COPD or asthma for adults ages 40 to 64. PQI-08 measures the number of inpatient hospitalizations for heart failure for adults age 18 to 64. PQI-15 measures inpatient hospitalizations for COPD or asthma for adults 18 to 39.

  Low value care. We use 5 claims-based measures from Charlesworth et al. (2016) to measure low value care. These measures are recommendations from CMS or the Choosing Wisely initiative, which aims to avoid unnecessary medical tests, treatments, and procedures. We selected these 5 measures as they had both a large number of qualifying diagnoses for the denominator and a high overall prevalence of low value care conditional on that diagnosis.

- **Denied claims.** In our administrative claims data, we observe the final payment status of each encounter reported by the Medicaid managed care plans. Since there is very minimal cost-sharing in New York Medicaid, these administrative denials represent the denial of claims submitted to Medicaid managed care plans by healthcare providers. We are unable to observe the reasons for denial in our data. Denials may occur for several reasons, including duplicate claims being submitted, claims submitted with errors, and claims submitted for unapproved services. We evaluate the role of denied claims (which are paid $0 in our data) by re-pricing each denied claim using the pricing regression described in Appendix Section C.5.

B.2 Differential Reporting?

In Section 4 we document substantial spread across plans in their causal impacts on spending and utilization. Are these findings, which are established via claims data, likely to be driven by differential reporting in which our so-called "low-spending" plans merely report incomplete claims data to the regulator? Several facts and patterns suggest the answer is no:

- **External Validation.** An evaluation by the Lewin Group indicated that the New York data was ready for use in research (Lewin Group, 2012). This is not the case for Medicaid encounter data from all states.

- **MLR Reports/OIG.** The Office of the Inspector General examined New York Medicaid managed care plans in 2012 and found that differences in medical loss ratios (MLRs) across plans spanned a 27 percentage point spread (68% to 95%) (OIG, 2015). Though not directly numerically comparable to our log spending difference estimates (and impossible to correlate with our data due to de-identification of individual plans in the OIG report), these numbers indicate significant heterogeneity across plans in spending relative to (risk-adjusted) capitation payments. This is consistent with our interpretation of the claims data as revealing plan spending differences. Importantly, the cost data entering the MLR calculation do not originate from the same systems as our claims data, so this offers an independent corroboration of plan heterogeneity in spending per beneficiary.

- **FFS Claims.** Some services are carved out of MMC plan responsibility and reimbursed directly by the state under a FFS arrangement. If plans differed in reporting but not in utilization, then FFS claims—in which the plans themselves have no reporting role—may be similar across plans. Figure A6 compares FFS spending and total spending across enrollees assigned to different plans. FFS spending differs across plans and in fact closely tracks the across-plan differences in total spending. This is consistent with there being true utilization reductions by low spending plans and complementarities between MMC and FFS utilization. (And complementarity, rather than substitution, appears likely, as we show in Section 4.2 that MMC is a blunt instrument, reducing nearly all types of utilization.)

- **Incentives.** Finally, we merely note that plans have an incentive to make their claims reporting complete. Future market-level capitation rates depend on past market-level claims.

C Identification and Robustness

C.1 IV Assumptions

We briefly discuss the IV assumptions. We begin by noting that the fixed effects that control for the month t × county c of enrollment (ϕ_{ct}) are important here, as the month × county determines the set of plans eligible to receive enrollees via auto assignment.

- **Independence.** The intention of the state Medicaid administrator was to randomly assign auto-assignees across the eligible plans. Figure 1 provides strong evidence that plan assignment is indeed as good as random, by examining correlation between pre-determined characteristics and plan of assignment. Each pre-determined baseline characteristic is regressed on the full set of ten indicators for beneficiaries' assigned plans. The figure shows p-values from the F-tests of the null hypothesis that the baseline characteristics do not differ significantly among beneficiaries assigned to different plans. Successful random assignment would tend to generate large p-values, indicating no significant relationship. These values tend to be large across the pre-determined characteristics. We also present p-values for the same tests for an equal-sized (random) sample of active choosers, showing important relationships between choices and observables in the population that was not auto-assigned. This exercise demonstrates that these variables do indeed capture relevant dimensions on which enrollees select across plans. We also present the individual plan effect estimates for auto-assignees and active-choosers in the figure, showing that while the active-chooser estimates of plan "effects" on predetermined characteristics are often substantial (indicating imbalance), the auto-assignee estimates are typically close to zero (indicating balance).

- **Relevance.** From Figure 1, assignment has a clear, quantitatively large effect on enrollment. Beneficiaries assigned to a plan spend more than ninety percent of the months in a Medicaid episode in the plan to which they were assigned. In other words, the first-stage effect of plan of assignment on plan of enrollment is near one for all plans. The first-stage F-stat is above 7,000.
- **Exclusion.** Here, the exclusion-restriction requires that plan of assignment influences outcomes like healthcare utilization only via plan of enrollment. That is a natural assumption in this context, in which the plan of enrollment is the vehicle through which healthcare is provided. Although it is impossible to rule out, for example, that assignment to some plan—as distinct from enrollment in that plan—causes the healthcare utilization outcomes we document, such an interpretation would be significantly at odds with the existing small experimental and quasi-experimental literature on health plan effects. Another potential violation of the exclusion-restriction could occur if plan of assignment caused attrition out of the observation sample. This would be the case if plan of assignment caused beneficiaries to exit the Medicaid system altogether (as opposed to exiting the plan of assignment or exiting the managed care program to enroll in FFS Medicaid). We rule out this possibility directly in the data, showing no differential attrition over our study window (see Figure 4).
- **Monotonicity.** Finally, the monotonicity assumption is supported by the near-completeness of take-up. The first-stage enrollment effects are nearly 1.0, leaving little room for defiers.

Separate from these issues of instrument validity, it is useful to consider the LATE identified here. Compliers are those whose plan of observed enrollment (from the econometrician's viewpoint) is affected by assignment. First and most importantly, the LATE here is likely to be close to the ATE for the quasi-experimental sample because of the size of the first-stage: Over ninety percent of the enrollee-months in a Medicaid episode are spent in the plan of assignment. At the limit of one hundred percent, the LATE converges to the ATE. There is a subtlety in this setting because there are ten first stage regressions, each instrumenting for enrollment in one of the ten plans. Conceptually, this opens the possibility that the LATEs are different across the first-stages. Compliers could, in principle, be different types of enrollees with different counterfactuals across the ten assignment possibilities. The strongest evidence against this possibility is that the first-stage coefficient is so similar and so close to 1.0 across all first stage regressions.

C.2 Plan Group IV Regressions

In Section 4.2 we describe an IV regression in which the regressors are plan groups: low- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high-spending (Plan D). The corresponding equation is:

$$Y_{ict} = \rho + \psi_{ct} + \nu X_{ict} + \gamma_{\text{Low}} \mathbf{1}[\widehat{\text{Low Plan}}_{ict}] + \gamma_{\text{High}} \mathbf{1}[\widehat{\text{High Plan}}_{ict}] + \mu_{ict} \tag{A.1}$$

Medium spending is the omitted category. The endogenous variables are indicators for enrollment in any plan in each set, and the instruments are indicators for *assignment* to any plan in each set, so there are two first-stage regressions:

$$\text{Low Plan}_{ict} = \phi_{1ct} + \delta_1 X_{ict} + \sum_{j \in \text{High,Low}} \lambda_{1j} \mathbf{1}[\text{Assigned j}_{\ ict}] + \eta_{1,ict}, \tag{A.2}$$

$$\text{High Plan}_{ict} = \phi_{2ct} + \delta_2 X_{ict} + \sum_{j \in \text{High,Low}} \lambda_{2j} \mathbf{1}[\text{Assigned j}_{\ ict}] + \eta_{2,ict}. \tag{A.3}$$

Results using this specification are presented in Figure 4 and Tables A13 and A14.

C.3 Short- vs. Long-run Effects, Disruption, and Alternate Construction of Spending Variables

As discussed in Section 2, for all analyses in the main text we restrict to the six months following plan assignment. This is due to the fact that few auto-assignees remain enrolled after the sixth month post-assignment. This can be seen in Figure A1. Panel (a) shows the full length of enrollment spells for auto-assignees. Panel (b) shows post-assignment enrollment. The modal beneficiary is enrolled in Medicaid for 12 months, though many are enrolled for less than 12 months. Focusing on post-assignment enrollment, over 30% of auto-assignees are enrolled for exactly 6 months. Only a few auto-assignees remain enrolled past 6 months. Because of this, in the main text, we focus all of our attention on the first 6 months of enrollment post-assignment.

Here, we investigate (to the extent possible) whether our key results hold up as we add additional months post-assignment. In Table A15 we report coefficient estimates for low- vs. medium-plan effects over different time periods. In column 1 we report our original results, where we restrict to the first 6 months post-assignment. In columns 2 and 3 we maintain the same sample of beneficiaries, but we allow post-assignment months 7-9 and 7-12 to enter the regression, respectively. In columns 4 and 5, we restrict to balanced panels of beneficiaries enrolled for at least 9 and at least 12 months, respectively, also restricting the observations to 9 months and 12 months post-assignment. In all cases, the coefficients on "high" and "low" are virtually unchanged, though statistical power decreases for the "high" coefficient. These results provide evidence that the main effects presented in Table 1 and Figure 4 persist into the longer run—for the minority of managed care enrollees that have these longer enrollment spells.

To understand the role of disruption, we examine the time-pattern of spending effects in Figure A2, which plots IV estimates separately for each post-assignment month. Plans are divided into medium- and low-spending groups. The figure shows whether and how spending effects differ for low-spending plans relative to medium-spending plans in each month. Some disruption is likely to occur in any plan, but if disruption is relatively larger in low spending plans, one might expect a greater dip in spending in the earliest post-assignment months. Panels (a) and (b) use our original sample, with Panel (a) showing only the first 6 months post-assignment (for which we have data for everyone in the original sample) and Panel (b) extending up to 12 months post-assignment (where we only have data for a subset of beneficiaries beyond 6 months). Panels (c) and (d) use new balanced samples of beneficiaries with at least 9 and at least 12 months of post-assignment enrollment, respectively, so that the patterns over time cannot be explained by a change in the composition of beneficiaries remaining enrolled in Medicaid. In all cases, effects do appear somewhat larger in the first months post-assignment, but they remain large throughout the post-assignment months we analyze.

In Table A6, we investigate the sensitivity of our estimates in the main estimation sample to pooling per-enrollee spending over the entire six-month spell, rather than examining month-by-month spending. One practical consequence is that there are fewer observations (now enrollee-spells, rather than enrollee-months) with zero spending. This change in the underlying distribution of the dependent variable leads to spending results that are numerically different in the log specification (though not in the Winsorized level specification), with the aggregated spending estimates generally being larger than the monthly estimates. The table nonetheless shows that all specification variations yield results that qualitatively track the main spending estimates.

C.4 Networks

We briefly explore networks as a mechanism for the spending, quality, and satisfaction gaps we estimate across plans. We start by discussing how we construct the measures of healthcare provider network breadth we use to assess the role of networks.

We measure network breadth as the share of simulated physician and hospital visits from a given zip code covered by each plan's network. To simulate physician and hospital visits, we use estimates from models of physician and hospital demand in Wallace (2019), which include a "hassle cost" for going to an out-of-network provider. The estimates from these models are used to simulate where Medicaid enrollees would seek care if every provider was in-network. Lastly, as in Ericson and Starc (2015), the "simulated visit shares" measure is a calculation of the share of simulated physician and hospital visits for Medicaid enrollees living in a given zip code that are covered by each plan's network.[5]

We assess whether provider network breadth mediates our causal plan differences in two ways. First, we re-estimate the plan-level spending and satisfaction results in our randomly-assigned sample but with controls for network breadth. Appendix Table A16 presents the results of this analysis. Columns 1 and 2 contain comparisons of high vs. medium and low vs. medium spending estimates with and without controls for network breadth. Adding controls for network breadth does not change the large estimated differences across plan spending groups.

Second, we plot our causal estimates of plan-level spending and willingness-to-stay (i.e., consumer satisfaction) against plan-level measures of network breadth. We measure plan-level network breadth as the average of the zip-level network breadth measures, weighted by the zip code-specific count of Medicaid beneficiaries assigned to each plan. We plot these relationships in Appendix Figure A7. Panel A plots our estimated plan-level spending effects against network breadth measured at the plan-level. The slope of the line of best fit is close to zero. The same pattern emerges when we compare our estimated plan-level willingness-to-stay effects against network breadth in Panel B. We do not find strong evidence of a relationship between our plan-level spending and satisfaction estimates and provider network breadth. This can be reconciled with evidence from Gruber and McKnight (2016) and Wallace (2019) that broader provider networks increase spending and satisfaction by noting that the complex set of tools that modern health insurers rely on to constrain spending may counteract the effects of broader networks. For example, one of the Medicaid managed care plans in our sample is vertically-integrated with the public, safety net hospitals in New York City. That plan operates a very narrow hospital network but, when enrollees are randomly-assigned to it, we observe high levels of spending. Hence, it is likely that it combines a narrow hospital network with a relatively lenient set of utilization management tools. Indeed, this multi-dimensional nature of the contract is the primary motivation for the strategy used by Wallace (2019) to separately identify the effect of networks from other dimensions of MMC plans.

C.5 Re-pricing of Claims

In Section 4.2, we re-priced all claims to a common set of reference prices. To construct our list of common reference prices, we begin by following Cooper et al. (2019) in estimating plan "effects" on

[5] Additional details on network breadth measure construction and summary statistics are available in Wallace (2019).

prices. The estimating equation is:

$$P_{djc} = \nu_d + \sum_{j=1}^{9} \Psi_j \mathbf{1}[\text{Plan j }] + \mu_{djc}, \tag{A.4}$$

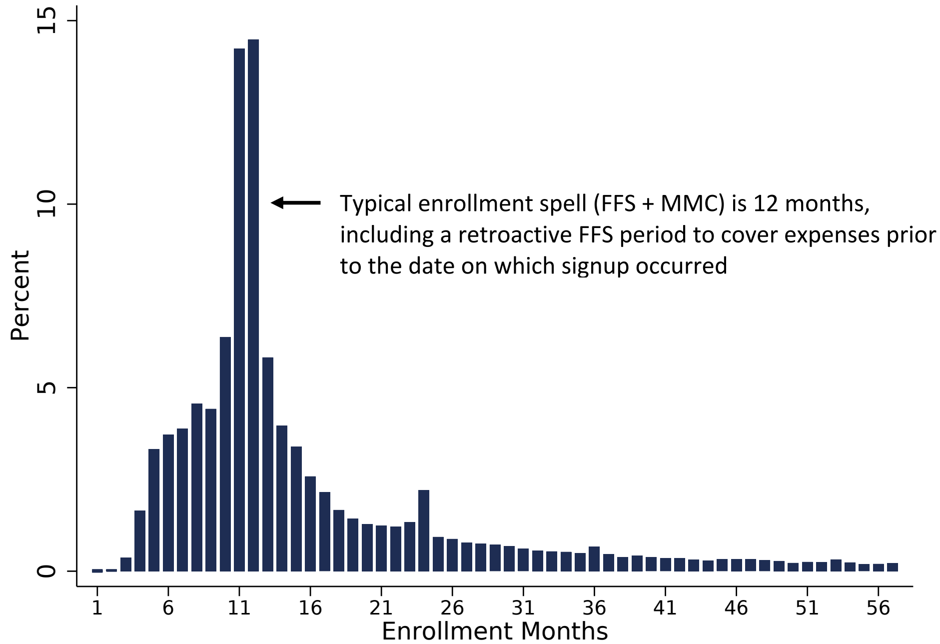
where P_{djc} indicates the log price paid by plan j for service d on individual claim record c in our data. Services d are comprised of DRGs for inpatient admissions and HCPCS for outpatient procedures. The regressors include service fixed effects (ν_d) and nine plan fixed effects (Ψ_j) that indicate the relative price level of each plan. If the data generating process underlying prices consisted of each plan determining prices as a constant-multiple markup for all services relative to some common index price for each service (such as the FFS Medicaid price), then Ψ_j would exactly recover that markup.

To reprice the claims, we use predicted values from this regression, assigning a common price across plans for each procedure. This common price is set to equal $e^{(\nu_d+\Psi_X)}$—the procedure fixed effect plus the plan effect from the omitted plan, de-logged. To establish the robustness of this price standardization method, in Figure A8 we plot the IV plan spending effects from this regression standardization against the IV plan spending effects based on the alternative price standardization method of setting all prices to the across-plan median. The two repricing methods yield nearly identical results.

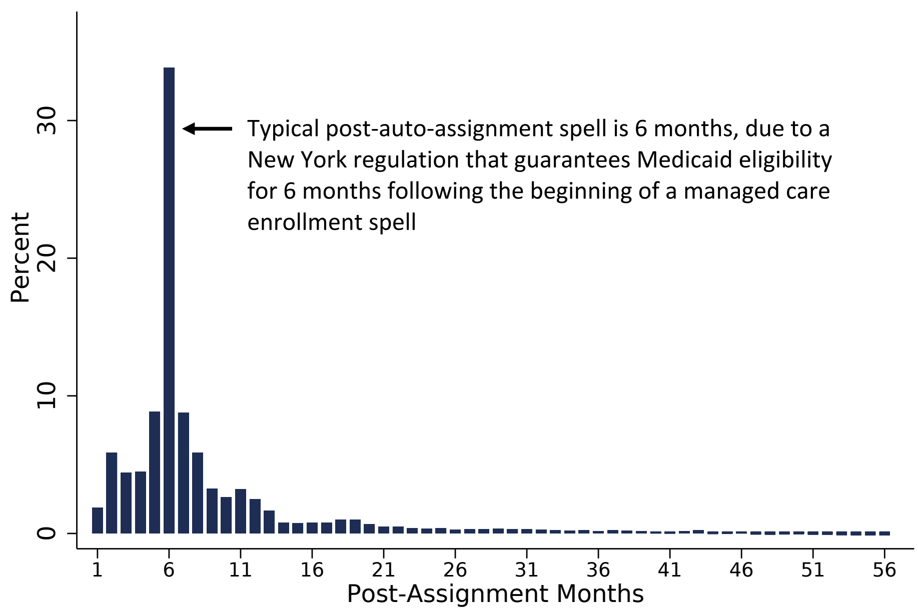
D Additional Figures and Tables

Appendix Figure A1: Enrollment Spell Lengths of Auto-Assignees

(a) Medicaid Enrollment (FFS plus MMC) Spell Lengths



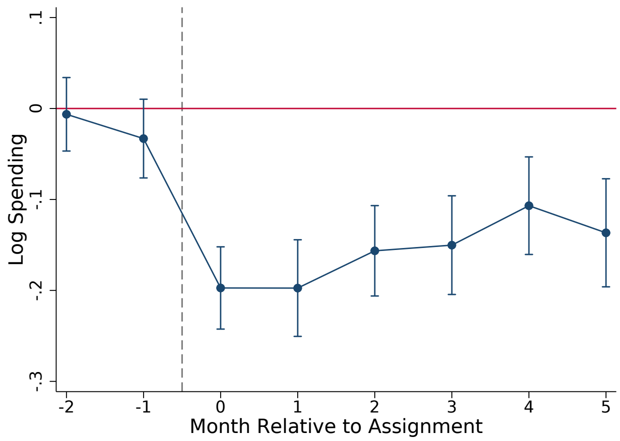
(b) Post-Auto-Assignment (MMC) Spell Lengths



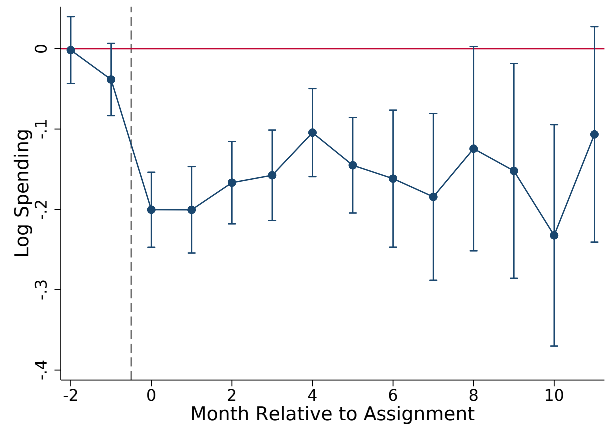
Note: Figure displays histograms of enrollment spells in our data for auto-assignees, prior to making sample restrictions based on enrollment length. The top panel shows the length of the overall Medicaid enrollment spell, which includes a fee-for-service (FFS) spell prior to assignment and a managed care (MMC) spell post-assignment. The bottom panel shows the length of the managed care (MMC) spell post-assignment. The typical post-assignment spell is 6 months due to a NY regulation that guarantees Medicaid eligibility for 6 months following the beginning of a managed care enrollment spell.

Appendix Figure A2: Event Study Difference-in-Differences: Effects By Time Since Plan Assignment

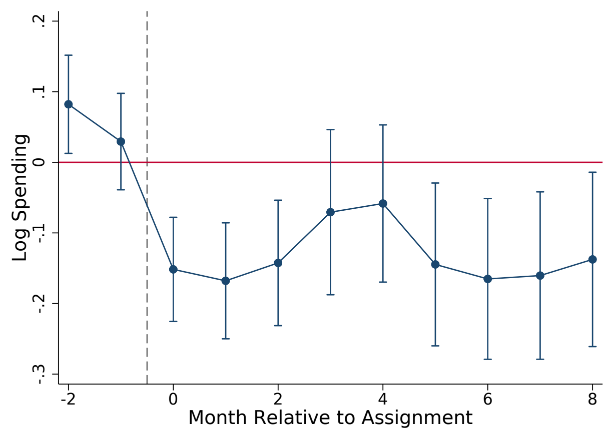
(a) Original sample



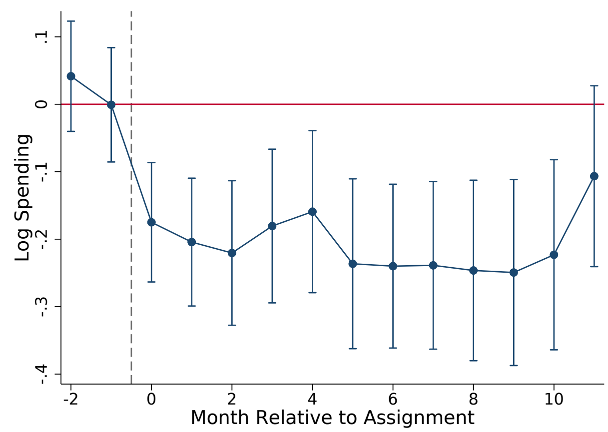
(b) Original sample extended to 12 months post-assignment



(c) Alternative 9 month post-assignment balanced sample



(d) Alternative 12 month post-assignment balanced sample



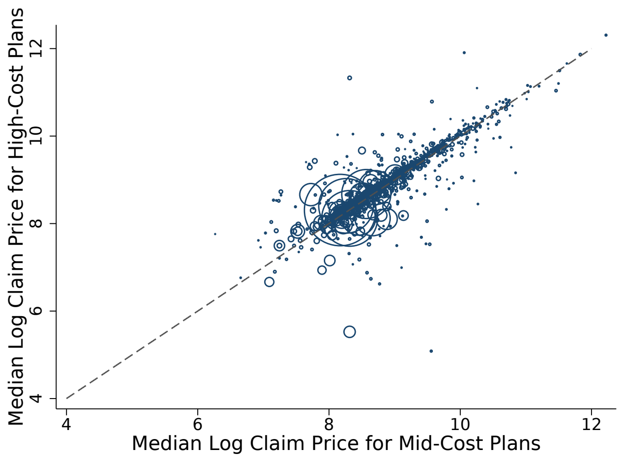
Note: Figure displays results in the spirit of difference-in-difference event studies showing the spending impacts of being assigned to a low- versus medium-spending plan. As in Table A13, we divide the ten plans into three sets: low- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high- (Plan D) spending plans. Medium-spending plans are the omitted category and results for low-spending plans are shown. Event time (τ) is along the horizontal axis with month zero corresponding to the first month post-assignment. Using a modification of the IV regression in Equation (4), each point is estimated from a separate regression (one for each τ) of the form:

$$\log(\text{Spending }+1)^{\tau}_{ict} = \alpha^{\tau} + \phi^{\tau}_{ct} + \delta^{\tau} X_{ict} + \lambda^{\tau}_{\text{low}} \mathbf{1}[\widehat{\text{Low Plan}}_{ict}] + \lambda^{\tau}_{\text{high}} \mathbf{1}[\widehat{\text{High Plan}}_{ict}] + \epsilon^{\tau}_{ict}.$$

We plot point estimates and 95% confidence intervals for $\lambda^{\tau}_{\text{low}}$. For the regressions corresponding to $\tau = -1$ and $\tau = -2$, we use a reduced form specification since enrollees are in FFS rather than any specific plan prior to assignment. The estimates show the (null) effect of a low plan relative to a medium plan on spending prior to the assignment occurring. For $\tau = -1$ and $\tau = -2$, spending is pre-randomization FFS spending, rather than post-assignment spending in managed care. No coefficients presented, including coefficients for $\tau = -1$ or $\tau = -2$, are normalized to zero. Panel (a) uses the main IV sample of auto-assignees and the main follow-up period of 6 months post-assignment. Panel (b) also uses the main IV sample of auto-assignees, but includes observations in months 7–12 post-assignment, if available for the beneficiary. This leads to an unbalanced sample over the event time window as many beneficiaries exit Medicaid after month 6. Panels (c) and (d) create new balanced samples that restrict to beneficiaries enrolled for at least 9 and at least 12 months, respectively, and restrict observations to the first 9 months and first 12 months post-assignment, respectively.

Appendix Figure A3: Price Comparisons Across High-, Medium-, and Low-Spending Plans

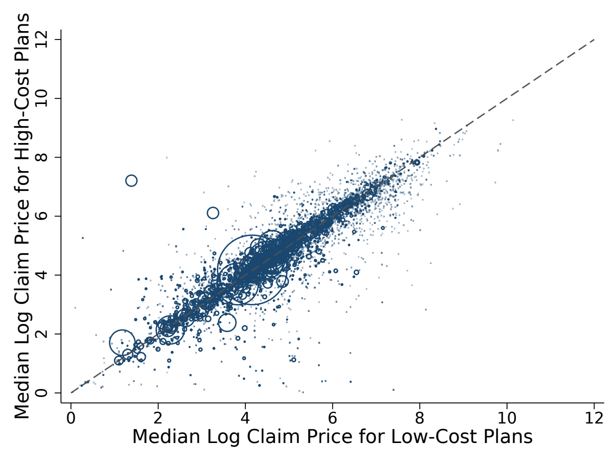
(a) Prices (DRGs) for High- vs. Medium-Spending Plans



(b) Prices (DRGs) for High- vs. Low-Spending Plans

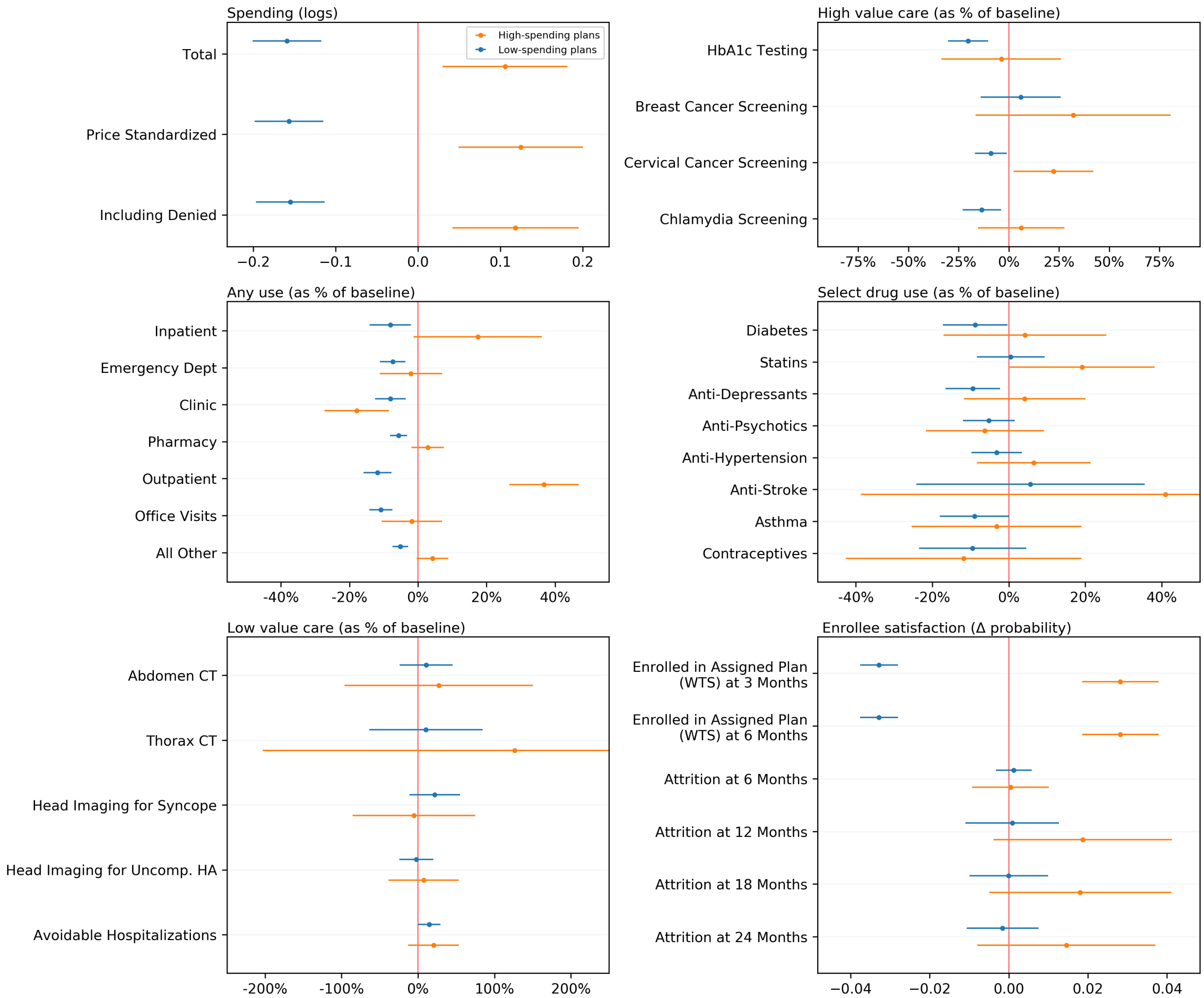
(c) Prices (HCPCS) for High- vs. Medium-Spending Plans

(d) Prices (HCPCS) for High- vs. Low-Spending Plans



Note: Figure compares prices for inpatient admissions and outpatient services between high-, medium-, and low-spending plans. We divide plans into high-, medium-, and low-spending groups as described in the text. Figure 3 shows analogous comparisons for low- versus medium-spending plans. Each circle represents a pricing unit: either a diagnosis-related group (DRG) in the case of inpatient prices or a Healthcare Common Procedural Coding System unit (HCPCS) in the case of outpatient prices. Marker size is proportional to frequency in our claims data. See Figure 3 for additional notes.

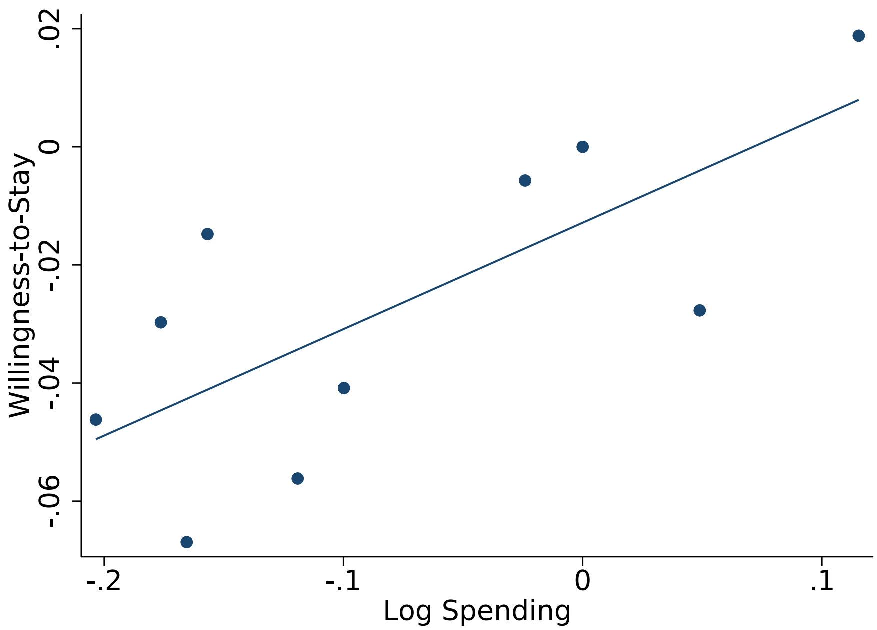
Appendix Figure A4: Extending the Figure 4 Results to the High-Spending Plan



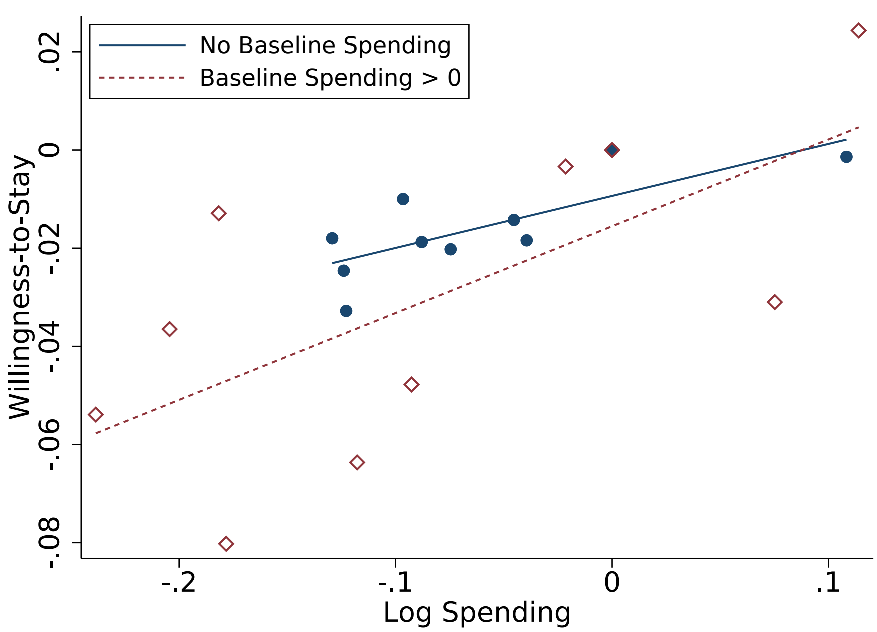
Note: Figure shows outcomes in low-spending plans and high-spending plans compared to medium-spending plans (omitted category) across various categories and service settings. See Figure 4 notes for additional detail.

Appendix Figure A5: Plan Satisfaction (WTS) versus Plan Spending Effects

(a) All Auto-Assignees

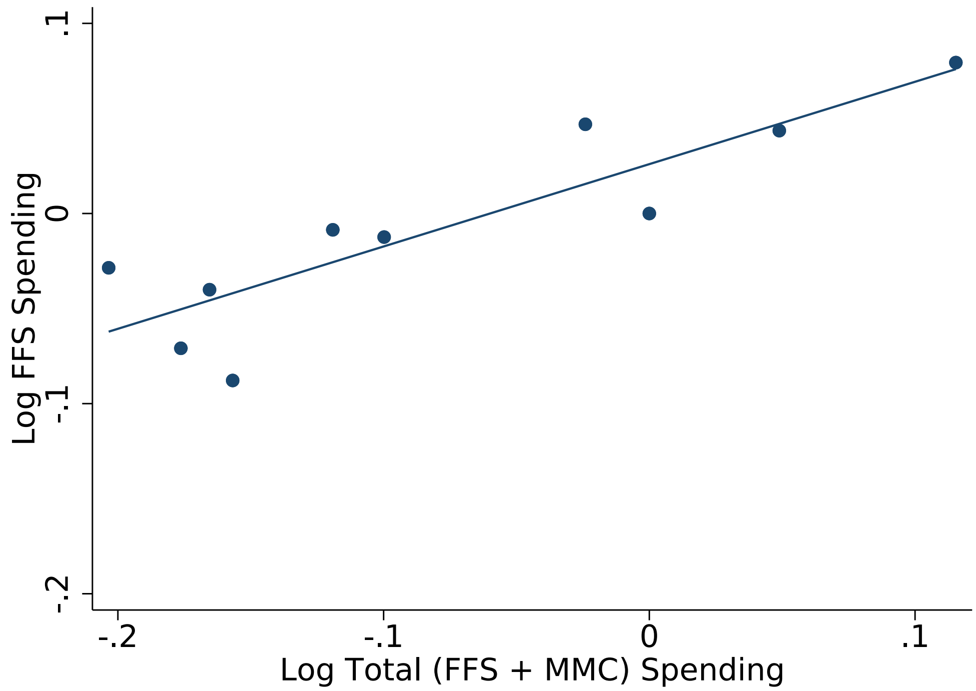


(b) Auto-Assignees Divided by Baseline (Pre-Assignment) Spending



Note: Figure shows the correspondence between willingness-to-stay (WTS) and IV plan spending effects. In the top panel, each plan corresponds to one point, with the coordinates corresponding to the coefficient estimates from Table 1. In the bottom panel, each plan corresponds to two points: The WTS and plan spending effects are estimated separately for enrollees with some spending during the baseline FFS period (prior to random assignment) and for enrollees with no spending during the baseline FFS period. The lines in each panel correspond to the OLS fit of the 10 points.

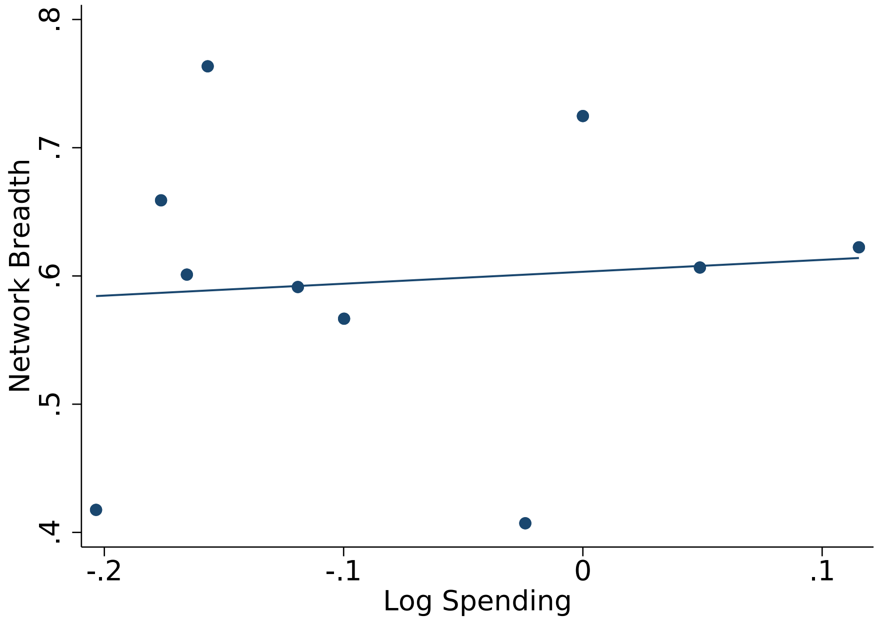
Appendix Figure A6: Carved-Out FFS Claims versus Total (FFS and MMC) Claims



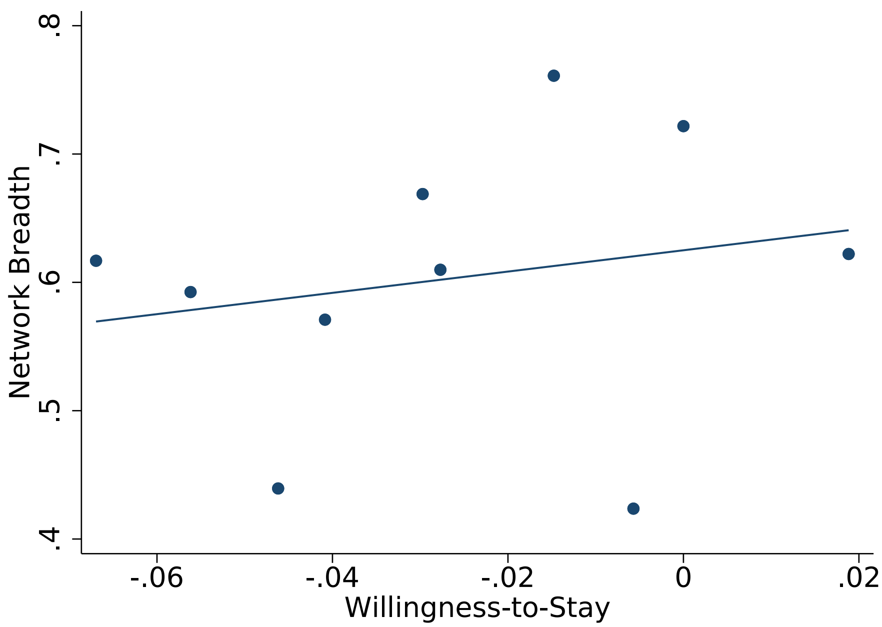
Note: Figure plots IV estimates of plan effects on carved-out FFS claims against plan effects on total (FFS plus MMC) claims. The sample is the main IV analysis sample. Carved-out FFS claims for MMC enrollees are paid and reported directly by the state, rather than by the plans despite occurring during MMC enrollment. Markers correspond to plans. The coefficients plotted along the horizontal axis are identical to those reported in Figure 3. Correlation between FFS claims and total claims is consistent with the joint hypothesis that low-spending plans affect spending across a broad set of services (including carved-out services) and that MMC claims data reveal true differences in utilization rather than merely differences in reporting (see Appendix Section B.2).

Appendix Figure A7: Association between Causal Plan Effects and Provider Network Breadth

(a) Spending vs Provider Network Breadth

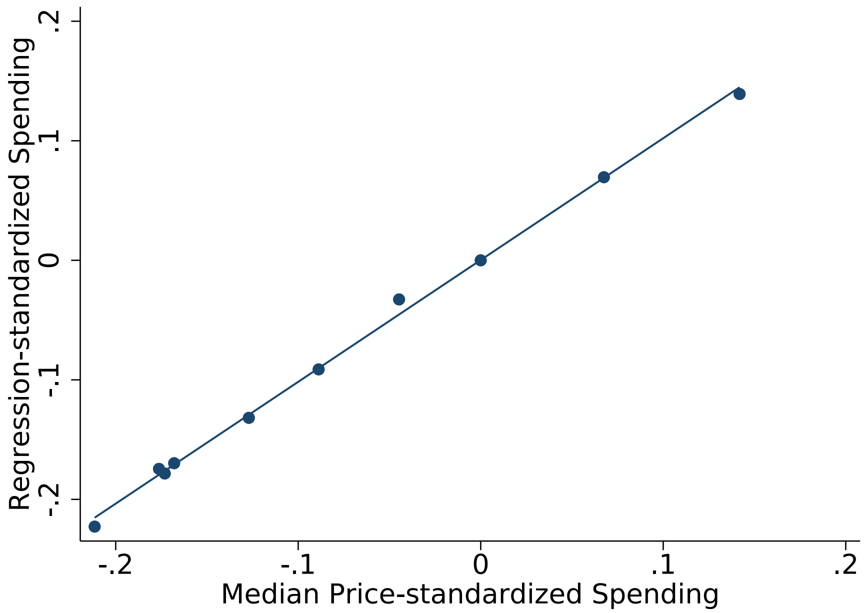


(b) Willingness-to-Stay vs Provider Network Breadth



Note: Figure displays the association between the main results of the paper—plan effects on healthcare spending and satisfaction—and plan-level measures of provider network breadth. Panel (a) plots IV coefficients corresponding to Eq. (4), where the dependent variable is log healthcare spending on the y-axis. Plan of enrollment is instrumented with plan of assignment. Coefficients are relative to the omitted plan, X. The x-axis contains the average network breadth for each plan, measured using the simulated visit shares measure. Panel (b) plots coefficients from the last column in Table 1, where the dependent variable is an indicator for whether an enrollee remained in their assigned plan at six months post-assignment. Coefficients are relative to the omitted plan, X. The x-axis contains the average network breadth for each plan, measured using the simulated visit shares measure. Appendix Section C.4 describes how the network breadth measure was constructed.

Appendix Figure A8: Plan Effects using Regression-standardized Spending vs. Median Price-standardized Spending



Note: Figure compares two different approaches to re-pricing claims prior to estimating IV plan spending coefficients. Along the horizontal axis, the common set of median prices are applied to each pricing unit (HCPCS and DRGs) for every plan and the IV plan effects on total spending are re-estimated. To generate coordinates along the vertical axis, plan price effects are estimated as coefficients via Equation A.4. (See Appendix C.5.) Then every price for every plan is replaced with the predicted prices for the omitted plan and the IV plan effects on total spending are re-estimated. The correspondence between the two sets of points indicates the extent to which the details of the repricing exercise matter.

Appendix Table A1: Summary Statistics

	Active Choosers		Auto-Assignees	
	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.587	0.492	0.401	0.490
White	0.338	0.473	0.272	0.445
Black	0.302	0.459	0.516	0.500
Age	34.3	12.7	35.7	12.7
Healthcare Spending:				
Total	464.35	2145.54	509.84	2825.82
Office Visits	66.05	580.01	21.52	165.30
Clinic	24.45	164.01	52.47	280.24
Inpatient	179.82	1704.67	220.30	2495.82
Outpatient	42.94	333.52	41.30	301.90
Emergency Dept.	10.00	79.18	15.60	98.93
Pharmacy	57.26	292.06	74.79	452.74
All Other	83.83	601.13	83.87	620.72
Drug Days Supply:				
Diabetes	1.64	11.11	1.11	8.69
Statins	1.33	7.75	0.83	5.79
Anti-Depressants	0.83	6.19	1.31	7.80
Anti-Psychotics	0.59	5.33	1.49	8.64
Anti-Hypertension	1.56	8.71	1.32	7.91
Anti-Stroke	0.08	1.98	0.10	2.14
Asthma	0.44	4.04	0.46	4.11
Contraceptives	0.59	5.18	0.25	3.28
High-Value Care:				
HbA1c Testing	0.0096	0.0973	0.0055	0.0739
Breast Cancer Screening	0.0051	0.0710	0.0015	0.0383
Cervical Cancer Screening	0.0245	0.1547	0.0073	0.0851
Chlamydia Screening	0.0142	0.1184	0.0066	0.0810
Low-Value Care:				
Abdomen CT	0.0006	0.0253	0.0003	0.0179
Head Imaging for Syncope	0.0003	0.0193	0.0004	0.0288
Head Imaging for Uncomp. HA	0.0019	0.0488	0.0019	0.0559
Thorax CT	0.0001	0.0100	0.0001	0.0092
Avoidable Hospitalizations	0.0015	0.0393	0.0054	0.0736
Observations	6066972		393576	

Note: Table presents summary statistics for our main analysis sample ("auto-assignees") and a comparison sample of Medicaid beneficiaries who made an active choice ("active choosers") and so were not included in the IV sample. Rows report means and standard deviations of the indicated characteristics. See Table A10 notes for a complete listing of the therapeutic classes included in each grouping of prescription drugs. See Appendix B for detailed descriptions of the low- and high-value care measures.

Appendix Table A2: First Stage Estimates: Plan of Assignment Predicts Plan of Enrollment

	(1) A	(2) B	(3) C	(4) D	(5) E	(6) F	(7) G	(8) H	(9) I
A	0.924**	0.001+	0.001	0.005**	0.000	0.012**	-0.000	0.001+	0.000
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.000)
B	0.000	0.906**	0.003**	0.012**	0.001	0.017**	-0.000	0.002**	0.001*
	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.000)
C	0.002	0.002*	0.940**	0.001	0.001	-0.002	0.001	0.000	0.000
	(0.001)	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.000)
D	0.000	-0.002	-0.001	0.955**	-0.002*	-0.005*	-0.002*	0.002*	-0.000
	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.002)	(0.001)	(0.001)	(0.000)
E	-0.003**	-0.001*	-0.001	0.004**	0.939**	0.010**	-0.003**	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.000)
F	0.001	0.001	0.000	0.006**	0.001	0.933**	0.002*	0.002**	0.001+
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.000)
G	0.000	0.001*	0.002**	0.009**	-0.000	0.013**	0.915**	0.002**	0.001+
	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)	(0.001)	(0.000)
H	0.001	0.001	0.001	0.008**	-0.001	0.001	0.001	0.932**	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.000)
I	0.001	-0.001	0.004	0.005+	0.001	0.001	0.002	0.002+	0.933**
	(0.002)	(0.001)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)	(0.006)
Observations	393576	393576	393576	393576	393576	393576	393576	393576	393576

Note: Table reports coefficients from the nine first stage regressions defined in Equation 3. In each regression, the outcome is a binary indicator for being enrolled in one of the ten plans. The right-hand-side variables of interest—the plan assignment instruments—are nine indicators for whether the individual was assigned to each of the plans. All regressions control for county × year × month-of-assignment and the count of months since plan assignment, both as indicators. Person level controls, as described in Table 1 are included as well in all columns. Standard errors in parentheses are clustered at the county × year × month-of-assignment level. This is the level at which the randomization operates. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Appendix Table A3: Balance in Predetermined Characteristics Across Plan of Assignment

	Auto-Assignee		Active Chooser	
	F-stat	P-Value	F-stat	P-Value
Female	1.2	0.28	617.1	0
Black	1.2	.29	608.6	0
SSI	1.0	0.46	142.5	0
Other	1.2	0.27	213.4	0
Dental	1.0	0.40	325.2	0
Transportation	0.9	0.54	50.2	0
Lab	1.0	0.43	266.8	0
Pharmacy	1.4	0.18	257.3	0
Inpatient, Non-delivery	0.7	0.68	202.7	0
Inpatient, Delivery	1.0	0.45	163.4	0
Emergency Dept	0.9	0.55	716.1	0
Specialist, Hospital	0.4	0.95	120.8	0
Specialist, Clinic	0.4	0.96	131.5	0
Specialist, Office	2.4	0.01	272.0	0
Primary Care, Hospital	0.7	0.69	901.9	0
Primary Care, Clinic	0.7	0.75	365.9	0
Primary Care, Office	0.7	0.75	70.9	0

Note: Table reports results from balance tests on the pre-determined characteristics of auto-assignees who are randomized to different plans, and of active choosers who selected different plans. These tabulated values are used in the plot in Figure 1. Pre-determined characteristics include demographics and healthcare utilization in FFS Medicaid prior to joining a managed care plan. Each managed care enrollee spent a pre-period (often a few months, once retroactive enrollment is included) enrolled in the FFS program prior to choosing or being assigned to a managed care plan. Two samples are used: the main IV analysis sample of auto-assignees (AA) and a same-sized random subsample of active choosers (AC), for comparison. Each pre-determined characteristic is regressed on the set of indicators for the assigned plan (for auto-assignees) or for the chosen plan (for active choosers). We report the p-values from F-tests that the plan effects in these regressions are jointly different from zero. Large p-values are consistent with random assignment. Small p-values indicate selection (endogenous sorting).

Appendix Table A4: No Differential Attrition Out of Medicaid Program Across Plan of Assignment

	(1) 6 Months	(2) 12 Months	(3) 18 Months	(4) 24 Months
A	0.002	0.008	0.003	-0.000
	(0.004)	(0.011)	(0.010)	(0.009)
B	-0.005	-0.014	-0.007	-0.011
	(0.004)	(0.011)	(0.009)	(0.008)
C	-0.002	-0.011	-0.020^{+}	-0.013
	(0.005)	(0.013)	(0.011)	(0.010)
D	-0.003	0.013	0.012	0.011
	(0.006)	(0.013)	(0.013)	(0.012)
E	-0.007	-0.013	-0.009	-0.009
	(0.004)	(0.008)	(0.009)	(0.008)
F	-0.002	-0.006	-0.008	-0.002
	(0.004)	(0.011)	(0.010)	(0.008)
G	-0.005	-0.001	0.006	0.006
	(0.004)	(0.010)	(0.009)	(0.008)
H	0.001	-0.018	-0.018	-0.013
	(0.005)	(0.012)	(0.011)	(0.009)
I	-0.002	-0.006	-0.017	-0.011
	(0.007)	(0.014)	(0.012)	(0.012)
Mean	.956	.350	.272	.199
Observations	34611	34611	34611	34611

Note: Table reports on the probability of continued enrollment in Medicaid—in any managed care plan or in fee-for-service—as a function of plan of assignment. The sample is restricted to enrollees auto-assigned to plans prior to January 2011, in order to allow a full 24 month run out and therefore keep a consistent sample across columns (i.e., to avoid censoring due to the end date of our data). Attrition out of the Medicaid program would imply attrition out of our data and sample. The table displays regression coefficients for plan of assignment, where coefficients are relative to the omitted plan (X). The dependent variables are indicators for continued enrollment at 6, 12, 18, and 24 months, as indicated. See Appendix A.3. Observations are enrollees. Standard errors in parentheses are clustered at the county $\times$ year $\times$ month-of-assignment level. This is the level at which the randomization operates. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$ $p < 0.01$.

Appendix Table A5: Alternative Specifications for Main IV Results: Monthly Spending

	(1) Log Total	(2) Inverse Hyperbolic Sine	(3) Winsorized	(4) Any Spending	(5) Log Std. Pay	(6) Winsorized Std. Pay
A	-0.100**	-0.109**	-22.972	-0.013*	-0.093**	-21.167
	(0.035)	(0.039)	(15.696)	(0.005)	(0.035)	(14.921)
B	-0.166**	-0.183**	-29.869^{+}	-0.025**	-0.157**	-30.494^{+}
	(0.037)	(0.041)	(17.567)	(0.006)	(0.037)	(16.536)
C	-0.157**	-0.172**	-40.246*	-0.022**	-0.175**	-54.256**
	(0.036)	(0.040)	(17.227)	(0.006)	(0.036)	(16.026)
D	0.115**	0.129**	64.052*	0.018**	0.130**	66.997*
	(0.041)	(0.045)	(25.948)	(0.007)	(0.041)	(25.960)
E	0.049	0.054	16.595	0.007	0.043	10.143
	(0.033)	(0.036)	(16.637)	(0.005)	(0.033)	(15.902)
F	-0.024	-0.029	-4.832	-0.007	-0.030	-17.892
	(0.031)	(0.034)	(17.644)	(0.005)	(0.031)	(16.749)
G	-0.119**	-0.133**	8.781	-0.021**	-0.127**	3.834
	(0.034)	(0.038)	(17.580)	(0.005)	(0.034)	(16.954)
H	-0.176**	-0.192**	-56.226**	-0.023**	-0.170**	-54.686**
	(0.038)	(0.043)	(17.694)	(0.006)	(0.039)	(16.889)
I	-0.203**	-0.228**	-3.496	-0.037**	-0.217**	-14.093
	(0.071)	(0.078)	(32.935)	(0.011)	(0.070)	(30.445)
Mean	2.09	2.33	445.89	0.35	2.09	426.91
Observations	393576	393576	393576	393576	393576	393576

Note: Table reports IV estimates of each plan's causal effect on utilization relative to an omitted plan (X), using the IV regression in Equation 4. The columns vary the parameterization of spending used as the dependent variable, as indicated in the column headers. For columns with price-standardized spending ("Std."), we first reprice all claims across all plans to a common set of prices and then re-estimate the IV specifications for plan effects on spending. The repricing follows the procedure used to create Figure 3 Panel c and is described in full detail in Appendix C.5. Winsorized outcomes are Winsorized above only, at the 99th percentile. "Any Spending" is a binary variable for the presence of any paid claim. All regressions control for county × year × month-of-assignment and the count of months since plan assignment, both as a set of indicators. Person level controls, as described in Table 1 are included as well in all columns. Observations are enrollee × months. Standard errors in parentheses are clustered at the county × year × month-of-assignment level. This is the level at which the randomization operates. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A6: Alternative Specifications for Main IV Results: Aggregate Spending

	(1) Log Total	(2) Inverse Hyperbolic Sine	(3) Winsorized	(4) Any Spending	(5) Log Std. Pay	(6) Winsorized Std. Pay
A	-0.258**	-0.277**	-161.743+	-0.077*	-0.237**	-174.301+
	(0.054)	(0.058)	(96.961)	(0.032)	(0.053)	(89.574)
B	-0.360**	-0.387**	-232.867*	-0.166**	-0.325**	-240.438*
	(0.058)	(0.063)	(108.258)	(0.037)	(0.058)	(100.474)
C	-0.227**	-0.243**	-286.975**	-0.133**	-0.254**	-373.076**
	(0.054)	(0.059)	(93.261)	(0.035)	(0.053)	(86.002)
D	0.162*	0.176*	435.338**	0.112**	0.195*	416.851**
	(0.077)	(0.083)	(160.107)	(0.039)	(0.077)	(152.929)
E	0.116*	0.124*	138.863	0.070*	0.108*	78.463
	(0.055)	(0.059)	(103.301)	(0.032)	(0.054)	(95.748)
F	-0.007	-0.007	-27.731	-0.042	-0.010	-113.543
	(0.054)	(0.059)	(106.276)	(0.028)	(0.054)	(95.308)
G	-0.251**	-0.273**	34.730	-0.125**	-0.258**	-3.372
	(0.054)	(0.058)	(100.130)	(0.031)	(0.053)	(94.443)
H	-0.330**	-0.353**	-307.603**	-0.136**	-0.308**	-332.770**
	(0.059)	(0.064)	(103.981)	(0.037)	(0.059)	(95.889)
I	-0.331**	-0.363**	-17.613	-0.218**	-0.354**	-33.346
	(0.111)	(0.121)	(196.497)	(0.064)	(0.111)	(183.211)
Mean	4.41	4.84	2738.26	2.09	4.40	2584.51
Observations	65596	65596	65596	65596	65596	65596

Note: Table reports IV estimates of each plan's causal effect on utilization relative to an omitted plan (X), using the IV regression in Equation 4. The columns vary the parameterization of spending used as the dependent variable, as indicated in the column headers. The difference here compared to Table A5 is that spending and utilization outcomes are totalled over the full six-month enrollment spell. The endogenous variables instrumented are the fraction of the enrollment spell spent in the indicated plan. Observations are enrollees. See Table A5 notes for additional details. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A7: Healthcare Spending for Active Chooser Sample

	(1) Unadjusted	(2) Weighted	(3) Risk Adjusted	(4) Risk Adjusted, Weighted	(5) Any Utilization	(6) Standardized	(7) Denied
A	-0.282**	-0.385**	-0.273**	-0.308**	-0.039**	-0.252**	-0.295**
	(0.009)	(0.043)	(0.015)	(0.048)	(0.002)	(0.015)	(0.014)
B	-0.545**	-0.779**	-0.449**	-0.532**	-0.060**	-0.389**	-0.376**
	(0.010)	(0.052)	(0.024)	(0.042)	(0.003)	(0.023)	(0.024)
C	0.049**	0.079+	-0.145**	-0.064	-0.014**	-0.143**	-0.188**
	(0.011)	(0.047)	(0.014)	(0.043)	(0.002)	(0.014)	(0.013)
D	0.199**	0.110**	-0.059**	-0.081*	0.000	-0.010	-0.014
	(0.008)	(0.037)	(0.009)	(0.033)	(0.001)	(0.009)	(0.009)
E	0.100**	0.111**	-0.055**	0.054+	-0.009**	-0.057**	-0.064**
	(0.009)	(0.038)	(0.013)	(0.033)	(0.002)	(0.013)	(0.013)
F	0.282**	0.217**	-0.102**	-0.045	-0.016**	-0.095**	-0.137**
	(0.008)	(0.037)	(0.012)	(0.032)	(0.002)	(0.012)	(0.012)
G	0.030**	-0.290**	-0.204**	-0.265**	-0.030**	-0.223**	-0.244**
	(0.011)	(0.046)	(0.014)	(0.032)	(0.002)	(0.013)	(0.013)
H	-0.053**	0.134**	-0.079**	0.052	0.003	-0.025	-0.039*
	(0.010)	(0.047)	(0.020)	(0.055)	(0.003)	(0.020)	(0.019)
I	-0.507**	-0.527**	-0.354**	-0.319**	-0.050**	-0.376**	-0.394**
	(0.013)	(0.063)	(0.017)	(0.058)	(0.003)	(0.018)	(0.018)
Mean	2.797	2.890	2.797	2.890	0.492	2.790	2.826
Observations	6066972	1673206	6066972	1673206	6066972	6066972	6066972

Note: Column 1 repeats the specification from Table 1, column 4. Column 2 reweights the active chooser sample to match the auto-assignee (IV) sample based on observable characteristics. Weights are set to equalize sizes of cells defined by the interactions of: deciles of FFS (prior to managed care enrollment) spending, sex, six age groups, five race groups, and each county $\times$ year $\times$ month tuple. Risk adjusted regressions include the following person-level controls: sex, 5 race categories, deciles of spending in FFS prior to MMC enrollment, and 47 age categories (single years from 18 to 64). All regressions control for county $\times$ year $\times$ month-of-assignment and the count of months since plan assignment/plan enrollment, both as indicators. Standard errors in parentheses are clustered at the county $\times$ year $\times$ month-of-assignment level. This is the level at which the randomization operates. See Table 1 notes for additional specification details. $^{+}\ p < 0.1$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$

Appendix Table A8: Main IV Results for Utilization by Category

	(1) Inpatient	(2) Emergency Dept	(3) Clinic	(4) Pharmacy	(5) Outpatient	(6) Office Visits	(7) All Other
High-Cost Plans	0.004^{+}	-0.001	-0.019^{**}	0.006	0.032^{**}	-0.002	0.009^{+}
	(0.002)	(0.003)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
Low-Cost Plans	-0.002^{**}	-0.004^{**}	-0.008^{**}	-0.013^{**}	-0.010^{**}	-0.010^{**}	-0.011^{**}
	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Dependent Mean	0.022	0.054	0.094	0.210	0.083	0.084	0.195
Observations	393576	393576	393576	393576	393576	393576	393576

Note: Table reports IV regression results for category or place of service, using a modification to the IV regression in Equation 4. The dependent variables, corresponding to the column headers, are binary variables for whether there was any use of the indicated category/place of service in the enrollee × month. To construct the plan group regressors, we divide the ten plans into three sets: low- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high- (Plan D) spending plans. Medium plans are the omitted category. The endogenous variables are indicators for enrollment in any plan in each set, and the instruments are indicators for assignment to any plan in each set. See Equation (A.1) in Appendix C.2. All regressions control for county × year × month-of-assignment and the count of months since plan assignment, both as sets of indicators. Person level controls, as described in Table 1 are included as well in all columns. Observations are enrollee × months. Standard errors in parentheses are clustered at the county × year × month-of-assignment level. This is the level at which the randomization operates. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A9: Main IV Results for High-Value Care Measures

	(1) HbA1c Testing	(2) Breast Cancer Screening	(3) Cervical Cancer Screening	(4) Chlamydia Screening
High-Cost Plans	-0.0002	0.0005	0.0018^{*}	0.0004
	(0.0010)	(0.0004)	(0.0008)	(0.0008)
Low-Cost Plans	-0.0013^{**}	0.0001	-0.0007^{*}	-0.0009^{**}
	(0.0003)	(0.0002)	(0.0003)	(0.0003)
Dependent Mean	0.0055	0.0015	0.0073	0.0066
Observations	393576	393576	393576	393576

Note: Table reports IV regression results for use of "high-value care," using a modification to the IV regression in Equation 4. The dependent variables, corresponding to the column headers, are binary variables for whether the indicated care was provided, conditional on the demographic and clinical qualifications that would warrant that care, in the given enrollee × month. See Appendix B for detailed descriptions of the inclusion criteria for each measure. Specification details follow Table A8. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A10: Main IV Results for Utilization of Select Drug Categories

	(1) Diabetes	(2) Statins	(3) Anti-Depressants	(4) Anti-Psychotics	(5) Anti-Hyper-tension	(6) Anti-Stroke	(7) Asthma	(8) Contra-ceptives
High-Cost Plans	0.001	0.005*	0.002	-0.003	0.002	0.001	-0.001	-0.001
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.001)	(0.002)	(0.001)
Low-Cost Plans	-0.002*	0.000	-0.004*	-0.002	-0.001	0.000	-0.002+	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Dependent Mean	0.022	0.024	0.034	0.038	0.033	0.003	0.016	0.008
Observations	393576	393576	393576	393576	393576	393576	393576	393576

Note: Table reports IV regression results for prescription drug fills, using a modification to the IV regression in Equation 4. The dependent variables, corresponding to the column headers, are binary variables for whether there was any use of the indicated drug group in the enrollee × month. Drug groups are supersets of REDBOOK therapeutic classes. Diabetes includes: Anti-diabetic agents, Sulfonylureas; Anti-diabetic agents, misc; Anti-diabetic agents, Insulins. Statins include: Anti-hyper-lipidemic Drugs. Anti-depressants include: Psychother, Anti-depressants. Anti-psychotics include: Psychother, Tranq/Antipsychotic; ASH, Benzodiazepines; Anticonvulsant, Benzodiazepine. Anti-hypertension includes: Cardiac, ACE Inhibitors; Cardiac, Beta Blockers; Cardiac, Alpha-Beta Blockers. Anti-stroke includes: Coag/Anticoag, Anticoagulants. Asthma/COPD includes: Adrenals Comb, NEC. Specification details follow Table A8. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A11: Main IV Results for Low-Value Care Measures

	(1) Abdomen CT	(2) Thorax CT	(3) Head Imaging for Syncope	(4) Head Imaging for Uncomp. HA	(5) Avoidable Hospitalizations
High-Cost Plans	0.00010	0.00011	-0.00002	0.00014	0.00105
	(0.00024)	(0.00015)	(0.00019)	(0.00046)	(0.00087)
Low-Cost Plans	0.00004	0.00001	0.00010	-0.00004	0.00076*
	(0.00007)	(0.00003)	(0.00008)	(0.00022)	(0.00039)
Dependent Mean	0.00033	0.00009	0.00049	0.00190	0.00544
Observations	393576	393576	393576	393576	393576

Note: Table reports IV regression results for use of "low-value care," using a modification to the IV regression in Equation 4. The dependent variables, corresponding to the column headers, are binary variables for whether the indicated category of low-value care was provided in the enrollee × month. See Appendix B for detailed descriptions of the low-value care measures. Specification details follow Table A8. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A12: Additional OLS Estimates of Plan Effects

	Active Choosers Only		Active Choosers and Auto-Assignees Pooled	
	Log Spending (Table 1)	Log Spending (Table 1)	Log Spending	Log Spending
Plan	(1)	(2)	(3)	(4)
A	-0.264**	-0.273**	-0.273**	-0.277**
	(0.024)	(0.015)	(0.023)	(0.015)
B	-0.564**	-0.449**	-0.559**	-0.449**
	(0.037)	(0.024)	(0.033)	(0.022)
C	0.046*	-0.145**	0.02	-0.155**
	(0.019)	(0.014)	(0.019)	(0.014)
D	0.216**	-0.059**	0.231**	-0.040**
	(0.013)	(0.009)	(0.012)	(0.009)
E	0.101**	-0.055**	0.115**	-0.032*
	(0.019)	(0.013)	(0.018)	(0.013)
F	0.290**	-0.101**	0.283**	-0.094**
	(0.019)	(0.012)	(0.018)	(0.012)
G	0.027	-0.204**	-0.038	-0.236**
	(0.023)	(0.014)	(0.023)	(0.013)
H	-0.056*	-0.079**	-0.075**	-0.096**
	(0.026)	(0.020)	(0.024)	(0.019)
I	-0.499**	-0.354**	-0.494**	-0.350**
	(0.022)	(0.017)	(0.021)	(0.016)
County x Year x Month FEs	X	X	X	X
Person-Level Controls		X		X
Obs: Enrollee X Months	6,067,014	6,067,014	6,460,590	6,460,590

Note: Table displays OLS results in which the dependent variable is the log of total plan spending in the enrollee-month. Columns 1 and 2 repeat specifications from Table 1. Columns 3 and 4 expand the sample to include the auto-assignees. The plan indicator regressors are defined as the plan initially chosen for the active choosers and as the plan initially assigned for the auto-assignees. See Table 1 for additional details on the specifications. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Appendix Table A13: Alternative Specifications for Main IV Results: Monthly Spending, Plan Groups

	(1) Log Total	(2) Inverse Hyperbolic Sine	(3) Winsorized	(4) Any Spending	(5) Log Std. Pay	(6) Winsorized Std. Pay
High-Cost Plans	0.106**	0.119**	57.778*	0.019**	0.125**	67.792**
	(0.038)	(0.042)	(25.247)	(0.006)	(0.038)	(24.995)
Low-Cost Plans	-0.159**	-0.175**	-31.499**	-0.022**	-0.157**	-28.517**
	(0.021)	(0.023)	(10.241)	(0.003)	(0.021)	(9.614)
Mean	2.09	2.33	445.89	0.35	2.09	426.91
Observations	393576	393576	393576	393576	393576	393576

Note: Table reports IV estimates of each plan grouping's causal effect on utilization, using a modification to the IV regression in Equation 4. We divide the ten plans into three sets: low- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high- (Plan D) spending plans. Medium plans are the omitted category. The endogenous variables are indicators for enrollment in any plan in each set, and the instruments are indicators for assignment to any plan in each set. See Eq. (A.1) in Appendix C.2. Specifications otherwise follow Table A5. See Table A5 notes for additional details. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A14: Alternative Specifications for Main IV: Aggregate Spending, Plan Groups

	(1) Log Total	(2) Inverse Hyperbolic Sine	(3) Winsorized	(4) Any Spending	(5) Log Std. Pay	(6) Winsorized Std. Pay
High-Cost Plans	0.132+	0.144+	401.850*	0.112**	0.171*	431.134**
	(0.070)	(0.076)	(158.638)	(0.036)	(0.070)	(150.202)
Low-Cost Plans	-0.315**	-0.339**	-201.393**	-0.133**	-0.304**	-186.526**
	(0.031)	(0.034)	(60.618)	(0.019)	(0.031)	(56.250)
Mean	4.41	4.84	2738.26	2.09	4.40	2584.51
N	65596	65596	65596	65596	65596	65596

Note: Table reports IV estimates of each plan grouping's causal effect on utilization, using a modification to the IV regression in Equation 4. We divide the ten plans into three sets: low- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high- (Plan D) spending plans. Medium plans are the omitted category. The endogenous variables are indicators for enrollment in any plan in each set, and the instruments are indicators for assignment to any plan in each set. See Eq. (A.1) in Appendix C.2. The difference here compared to Table A13 is that spending and utilization outcomes are totalled over the full six-month enrollment spell. Specifications otherwise follow Tables A6 and A13. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A15: IV Results Using Various Post-Assignment Observation Windows

	(1) 6 Months Balanced	(2) 6 Months Extended to 9	(3) 6 Months Extended to 12	(4) 9 Month Balanced	(5) 12 Months Balanced
High-Cost Plans	0.106**	0.109**	0.118**	0.108	0.106
	(2.75)	(2.69)	(2.82)	(1.23)	(1.13)
Low-Cost Plans	-0.159***	-0.153***	-0.152***	-0.129***	-0.182***
	(-7.50)	(-6.94)	(-6.48)	(-3.59)	(-4.70)
Observations	393576	492492	557316	237348	221574

Note: Table reports IV estimates of each plan grouping's causal effect on utilization, using a modification to the IV regression in Equation 4. We divide the ten plans into three sets: low- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high- (Plan D) spending plans. Medium plans are the omitted category. The first column reproduces column 1 from A13, which includes only the first six months post-assignment. See Table A13 for additional specification detail. Columns 2 and 3 maintain the same sample of enrollees as column 1, but include observations in months 7–9 and 7–12 post-assignment, respectively, in the regression. This leads to an unbalanced panel as many beneficiaries exit Medicaid after month 6. Columns 4 and 5 restrict to balanced panels of beneficiaries enrolled for at least 9 and at least 12 months, respectively, and restrict observations to the first 9 months and first 12 months post-assignment, respectively. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A16: IV Results With and Without Plan-by-Zip Controls for Network Breadth

	Total Spending		Willingness-to-Stay	
	W/o Ntwk Ctrl	W/ Ntwk Ctrls	W/o Ntwk Ctrl	W/ Ntwk Ctrl
High-Cost Plans	0.110**	0.100*	0.020**	0.017**
	(0.039)	(0.039)	(0.003)	(0.003)
Low-Cost Plans	-0.157**	-0.162**	-0.022**	-0.023**
	(0.021)	(0.021)	(0.002)	(0.002)
Observations	393576	393576	393576	393576

Note: Table reports IV estimates of each plan's causal effect on utilization relative to an omitted plan (X), using a modified version of the IV regression in Equation 4. In addition to instrumenting for plan with plan of assignment, we also instrument for network breadth (which varies at the plan-by-ZIP code level) using the network breadth of plan of assignment. Observations are enrollee $\times$ months. The dependent variable is log spending, as in the main specification in Table 1. All regressions control for county $\times$ year $\times$ month-of-assignment and the count of months since plan assignment, both as indicators. Person level controls, as described in Table 1 are included as well in all columns. Observations are enrollee $\times$ months. Standard errors in parentheses are clustered at the county $\times$ year $\times$ month-of-assignment level. This is the level at which the randomization operates. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.