NBER WORKING PAPER SERIES

WHAT DIFFERENCE DOES A HEALTH PLAN MAKE? EVIDENCE FROM RANDOM PLAN ASSIGNMENT IN MEDICAID

Michael Geruso
Timothy J. Layton
Jacob Wallace

Working Paper 27762
http://www.nber.org/papers/w2776

NATIONAL BUREAU FO ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2020, Revised October 2021

Previously circulated as “Are All Managed Care Plans Created Equal? Evidence from Random Plan Assignment in Medicaid.” We thank Zarek Brot-Goldberg, Marika Cabral, Michael Chernew, David Cutler, Joshua Gottlieb, Ben Handel, Robert Kaestner, Jon Kolstad, Neale Mahoney, David Meltzer, 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. Thanks to Orin Hassan, Anthony Lollo Jr., Julia Yates, and Philip Valtadoros for stellar research assistance. 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. 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, any funder, or the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w27762.ack

What Difference Does a Health Plan Make? Evidence from Random Plan Assignment in Medicaid
Michael Geruso, Timothy J. Layton, and Jacob Wallace
NBER Working Paper No. 27762
August 2020, Revised Octobe 2021
JEL No. H75,I11,I13

ABSTRACT

Exploiting the random assignment of Medicaid beneficiaries to managed care plans, we find substantial plan-specific spending effects despite plans having identical cost sharing. Enrollment in the lowest-spending plan generates 30% lower spending—driven by differences in quantity—relative to enrollment in the highest-spending plan. Rather than reducing “wasteful” spending, lower-spending plans broadly reduce medical service provision—including the provision of low-cost, high-value care—and worsen beneficiary satisfaction and health. Consumer demand follows spending: a 10 percent increase in plan-specific spending is associated with a 40 percent increase in market share. These facts have implications for the government’s contracting problem and program cost growth.

Michael Geruso
University of Texas at Austin
Department of Economics
1 University Station C3100
Austin, TX 78712
and NBER
mike.geruso@austin.utexas.edu

Timothy J. Layton
Harvard Medical School
Department of Health Care Policy
180 Longwood Avenue
Boston, MA 02115
and NBER
layton@hcp.med.harvard.edu

Jacob Wallace
Yale School of Public Health
Department of Health Policy and Management
60 College Street
New Haven, CT 06520
jacob.wallace@yale.edu

1 Introduction

There are perhaps no more pressing issues in US healthcare than how to divide the provision of health insurance across public and private sectors and how to constrain healthcare spending growth. Motivated by the goal of cost control and often undergirded by a belief that choice and competition between private plans can best satisfy heterogeneous consumer preferences, managed competition between private health plans has become the dominant policy choice for social health insurance provision in the United States (Gruber, 2017).

In Medicaid, which is the largest health insurer in the United States, nearly three-quarters of enrollees are now covered by private Medicaid managed care (MMC) plans (Figure 1).[1] The case for privatization in Medicaid often centers on cost control, but the evidence is mixed on whether managed care has led to a more efficient use of resources in the Medicaid program.[2] Tightly bound with this puzzle of rising program costs despite privatization is the unresolved question of what effect private health plans *can* exert on the medical spending of their enrollees. Given the widely documented selection issues in insurance markets, a reasonable hypothesis is that lower-spending plans merely attract lower-spending enrollees.

The issues of cost control and the question of whether and how plans impact enrollee healthcare consumption are linked through consumer choice: An important feature of publicly-subsidized health insurance markets is that beneficiaries choose among a set of differentiated health plans, with much of the differentiation now occurring on dimensions unrelated to demand-side cost-sharing. Nowhere is this more pronounced than in Medicaid. Because consumer cost-sharing (e.g., via copays and deductibles) is generally prohibited in Medicaid,[3] MMC plans must rely on other plan features to control cost and differentiate themselves to consumers. However, compared to the substantial academic literature on the impact of demand-side cost sharing, relatively little is known about whether and to what extent these supply-side (i.e., non-cost-sharing) tools are effective—despite their near-universal adoption by public and private insurance plans.

[1]In 2017, 54 million Medicaid beneficiaries (69%) and 19 million Medicare beneficiaries (33%) were enrolled in private managed care plans (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 insurers. By 2018, Medicaid spending comprised nearly 30% of spending from from all sources (Figure 1).

[2]There is evidence of savings for some categories of care (Dranove, Ody and Starc, 2017), but most of the literature shows mixed effects on program costs (Duggan and Hayford, 2013; Marton, Yelowitz and Talbert, 2014; Layton et al., 2019; Layton and Politzer, 2021).

[3]For evidence on the impact of demand-side cost sharing, see, for example, Manning et al. (1987), Aron-Dine, Einav and Finkelstein (2013), or Brot-Goldberg et al. (2017).

In this paper, we examine three interrelated questions. In the first part, we ask whether private managed care plans can affect patient healthcare spending (rather than merely attract high- or low-spending patients) without exposing consumers to cost-sharing. Second, we assess how spending reductions are achieved by managed care plans—and what trade-offs the savings entail. And, third, we ask whether competitive forces and consumer choice within Medicaid allocate beneficiaries to plans that efficiently constrain healthcare spending. The answers to these questions are of first-order importance in the study and practical management of publicly-subsidized private markets (in Medicaid and other settings). For example, if demand favors plans that cause both spending and clinical quality to *increase*, then, absent intervention from regulators, consumer choice will lead to higher spending and quality at the market level. Further, demand-driven competition will encourage plans to increase spending and quality, eventually leading to higher program costs and higher quality healthcare in equilibrium. On the other hand, if demand favors plans that cause spending to *decrease* (e.g., because plans offer more efficient care via effective preventive treatments, or because plans reinvest savings into effective marketing), then consumer choice will lead to lower spending at the market level, and plans will be encouraged to drive spending down. In this way, our research question echoes recent work on the allocative efficiency of the US healthcare system (Chandra et al., 2016; Chandra and Staiger, 2020).

To investigate, we leverage the random assignment of nearly 70,000 beneficiaries to Medicaid Managed Care (MMC) plans from 2008 to 2012. The setting for our natural experiment is New York City, the second-largest MMC market in the United States, where ten plans competed for enrollees during our study period. Like many state Medicaid programs, beneficiaries in New York who did not actively choose a plan within a designated choice period were *randomly assigned* to one (a process known as "auto-assignment"), allowing us to estimate causal plan differences in healthcare spending and patient outcomes in an IV framework. The key identification challenge we overcome—the endogenous sorting of beneficiaries across plans (see, e.g., Geruso and Layton, 2017)—parallels the difficulty of overcoming selection bias 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); neighborhood effects (Finkelstein, Gentzkow and Williams, 2016, 2019; Chetty, Hendren and Katz, 2016; Chetty and Hendren, 2018*a*,*b*); and teacher and school effects (Chetty, Friedman and Rockoff, 2014*a*,*b*; Angrist et al., 2016, 2017).

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.[4] This finding is, in itself, a striking new fact. To put this result in context, a 30% difference in total mean spending was the difference in the RAND health insurance experiment between the 0% and the 95% coinsurance arms (Manning et al., 1987). These results reveal that (at least some) insurers can significantly constrain healthcare spending, even in the absence of *any* demand-side cost-sharing (deductibles, coinsurance, and copayments). Comparing our IV estimates of plan spending effects based on random assignment to risk-adjusted observational measures reveals that they are correlated, but the risk-adjusted measures tend to overstate the causal differences in spending across plans.[5]

If lower negotiated provider prices accounted for the savings in low-spending plans, then spending reductions could have minimal effects on consumer well-being (being instead a transfer from providers to plans and ultimately to the public program, as we discuss below). However, we find that unlike in fully private health insurance 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 publicly-funded insurance setting. Instead, lower-spending plans—disproportionately for-profit entities—constrain the *quantity* of healthcare goods and services received by program beneficiaries, particularly on the extensive margin. We find that enrolling in the lowest-spending plan reduces a beneficiary's probability of receiving *any* care in a given month by about 5 percentage points (or 16 percent) relative to the highest spending plan.

The lower real resource use we document in low-spending plans suggests the possibility of a material trade-off, in which these plans restrict access to services, technologies, or providers valued by enrollees. In contrast, if lower-spending plans control cost by keeping beneficiaries healthy or better coordinating their care, consumers may be better off in these plans. (This is the positive case often made in favor of managed competition.) To assess this, we examine the types of services for which

[4]This spending gap does not fade over time, implying a persistent spending difference rather than merely a differential disruption of care in lower-spending plans. The magnitude of this finding is similar to that reported in contemporaneous work on spending variation between commercial health plans (Handel et al., 2019).

[5]This is consistent with classic adverse selection, wherein plans that do less to constrain spending—i.e., plans that provide more care—attract and retain sicker patients. This fact suggests that using observational measures of spending and quality to reward or penalize plans—a widespread practice—may inadvertently reward selection. Ordinal ranking, on the other hand, is largely-preserved, suggesting that policies based on relative spending or quality of plans in a market may be only somewhat affected by selection.

plans matter. We show that cost savings in the lower-spending plans are driven by broad-based reductions in care provided, including lower utilization of inpatient and outpatient care and prescription drugs.[6] We further establish that lower-spending plans are not merely cutting low value services (e.g., imaging for an uncomplicated headache) and promoting high value services (e.g., statins to control cholesterol).[7] Instead, managed care tools used by the lower-spending plans to constrain cost are blunt: Enrollees in the lower-spending plans used fewer of both low and high value services and were more likely to be hospitalized for avoidable reasons. An important implication of these findings is that—somewhat contrary to popular myth in the broader healthcare landscape—lower-spending plans are not achieving savings by keeping people healthy. They are restricting access to a broad set of services with potentially harmful health consequences.

Beneficiaries may or may not highly value the plan attributes reflected in these clinical measures. To build a more complete characterization of consumer well-being, we generate a novel revealed preference measure that uses the same identifying variation that identifies our plan effect estimates. The key insight is that beneficiaries' willingness to continue to comply with the random assignment reveals important information—their plan preferences post-assignment. While imperfect compliance poses no problem for identification in our IV framework, it does create an opportunity for identifying revealed preference. Using our measure of experienced utility, we show that lower-spending plans are significantly more likely than higher-spending plans (71%) to lose auto-assignees due to noncompliance. This suggests a real trade-off between spending and beneficiary satisfaction, a supply-side analog to the trade-off between risk protection and moral hazard inherent in the use of demand-side cost-sharing.

We conclude with an analysis of whether choice and competition in this setting lead beneficiaries to plans that effectively constrain spending, consistent with the positions of policymakers who advocate for the transition to private provision. What matters for the larger question of whether managed care can reduce spending in aggregate is the interaction of plan spending effects and enrollment flows among the overall population, including the active choosers not used in our IV analysis. There are many reasons to doubt that enrollment flows necessarily follow clinical measures of plan quality,

[6] These findings bear some resemblance to evidence from Curto et al. (2017) that, relative to Traditional Medicare, private Medicare Advantage plans generate lower health care spending primarily via broad-based reductions in utilization.

[7] We follow Brot-Goldberg et al. (2017) in defining and examining enrollees' use of high-value and low-value services (Schwartz et al., 2014). And we examine drugs and preventive services aimed at improving population health (Chernew, Schwartz and Fendrick, 2015).

given the type of choice frictions and imperfections often documented in this domain (e.g., Handel and Kolstad, 2015; Abaluck et al., 2021). Additionally, in Medicaid there are reasons to doubt that beneficiaries will flock to more efficient plans that are able to constrain spending, as plans have limited ability to pass savings back to beneficiaries in forms that beneficiaries value most, such as cash via lower premiums (as there are no premiums) and additional supplemental benefits (which are typically not allowed).[8] This differs, for example, from managed competition in Medicare (Song, Landrum and Chernew, 2013; Duggan, Starc and Vabson, 2016; Cabral, Geruso and Mahoney, 2018; Curto et al., 2021). For these reasons, it is unclear *ex ante* what types of plans beneficiaries will prefer and thus what types of plans this market will reward.

We study this question by observing beneficiaries making active plan choices. We find that demand follows spending. Health plans with 10% higher spending on healthcare among the randomly-assigned enrollees have a 4.1 percentage point (41 percent) higher market share among enrollees making active choices, echoing evidence that consumer demand follows quality in the hospital sector (Chandra et al., 2016). We further show that this pattern of demand following spending holds when examining both the origin and destination plans among auto-assignees who switch, when examining the initial choices of active chooser beneficiaries, and when examining the subsequent choices of active chooser beneficiaries who switch from their initial choice. Plan choices *do not* align with publicly-reported plan quality ratings, the one piece of information about plans provided to beneficiaries by the state at the time of choice. Instead, demand seems primarily tied to the ability to use care, and thus to higher levels of healthcare spending.

These findings regarding the correlation between demand and plan spending are critical for understanding the economics of Medicaid managed care and other public programs involving consumer choice among a set of government contractors. In particular, consumer-demand-driven competition in this setting will tend to drive up overall program spending, rather than reducing it. This is because program spending is tied to plan spending.[9] These allocation effects are non-trivial: If ac-

[8] Due to the lack of premiums, competition among plans in Medicaid managed care bears some resemblance to other markets with administratively set prices (e.g., hospital competition in the Medicare program), wherein firms compete for enrollees via non-price (i.e., quality) means (Gaynor and Town, 2011; Garthwaite, Ody and Starc, 2020).

[9] This is because plan payment rates are determined via average historical costs in practice. While managed care plans are capitated and the residual claimants of the healthcare spending of their enrollees, their reduced spending does actually decrease costs for the state. This occurs via the mechanism by which plan payment rates are determined. This mechanism involves trending past spending forward, resulting in spending reductions today causing lower plan payment rates tomorrow. This link between plan spending and state spending is highly salient to state policymakers, as evidenced by a frequent focus on keeping managed care plan spending down. We discuss this link in detail in Section 6.1.

tive chooser beneficiaries were randomly reallocated to be evenly distributed across plans, spending would be reduced by $559 million per year, or around 4% of total Medicaid spending on MMC in NYC. The facts we document highlight scope for policy remedies aimed at cost reductions, including smarter auto-assignment defaults (Wallace, 2020). Our results also imply that a state's choice of which managed care plans to contract with is not an innocuous one. In our setting, if the state removed the four highest spending plans from the market via a more managed procurement process, spending would decline by $1.4 billion per year or 10% of total NY Medicaid spending on MMC in NYC. The trade-off for that procurement decision would be declines in utilization/access, beneficiary satisfaction, and beneficiary health outcomes (as our IV estimation documents).

This paper contributes to a nascent literature on the effects of health plans in settings where plans differ on 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. (2019). 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 moral hazard, adding to a smaller recent literature concerned with these features (see, e.g., Curto et al., 2017; Layton et al., 2019). Consistent with Garthwaite and Notowidigdo (2019), we find substantial *causal* heterogeneity across plans in spending and utilization that arises without any differences in consumer cost-sharing exposure. Thus, significantly constraining healthcare spending does not require exposing consumers to out of pocket spending. In this way managed care circumvents the classic trade-off between financial risk protection and moral hazard noted by Zeckhauser (1970) and Pauly (1974).

Our findings also complement and extend an important literature dating 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, reducing the use of low- and high-value services alike (Brot-Goldberg et al., 2017). These findings sparked interest in whether managed care tools could better target inefficient utilization and manage the care of 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. They lead to broad reductions in utilization, limiting both high- and low-value care rather than targeting "waste."[10] Our results do not rule out the possibility that managed care tools could be used to efficiently ration and target healthcare products or services, but they do provide a well-identified and important data point on the "bluntness" of supply-side restrictions in practice.

The rest of the article proceeds as follows. Section 2 describes our empirical setting and data. Section 3 presents our empirical framework. Section 4 presents our main plan effect estimates for healthcare spending. In Section 5 we decompose the plan spending effects into price and quantity, and assess their correlation with causal estimates of plan effects on clinical quality and consumer satisfaction. Section 6 discusses the implications of our results for the economics of Medicaid managed care. Section 7 concludes.

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.[11] 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 the five counties comprising New York City, where enrollment in managed care is mandatory and which contains about two-thirds of the state's Medicaid 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 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. Critically, the 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

[10]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 negotiating lower provider prices.

[11]See Appendix A for additional detail.

whether beneficiaries remained in their assigned plans. We describe auto-assignment (our identifying variation) in the next section.

The claims data used to assess plan impacts on healthcare spending include information on providers, transaction prices, procedures, and quantities. All managed care plans are required to submit standardized encounter data for the services they provide, and the NYSDOH has linked these data to their own administrative records for claims paid directly by the state through the FFS program. Thus, the assembled data (at the enrollee-by-encounter level) contain beneficiary-level demographic and enrollment data, plan-reported claims-level data for each beneficiary while in an MCO, NYSDOH-generated claims-level data for FFS services prior to MCO enrollment, and NYSDOH-generated claims-level data for FFS services carved out of MCO responsibility during MCO enrollment.

In principle, the quality of managed care encounter data reported by MCOs may vary across markets and across plans within a market. For example, nationally-aggregated Medicaid managed care encounter data that is filtered through the Medicaid Analytic eXtract (MAX) is known to have quality problems for some states (though not New York in our time period; see Byrd, Dodd et al. (2015)), and may discard information that is idiosyncratic to a particular state or time period. It is important to understand that our data come directly from the NYSDOH and that New York during our sample period is a high-quality outlier in terms of MCO claims validation Lewin Group, 2012.

Using this data, we construct several beneficiary-month level outcomes:

Healthcare use, prices, and spending. We observe all services paid for by the managed care plans and by fee-for-service Medicaid. 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 on a variety of baseline characteristics. When we report total enrollee spending in managed care, we add together the components paid by the MCO plan as well the services carved out from managed care financial responsibility and paid for via FFS by the state.

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 as well as the number of 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 older 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 5.3, we measure willingness-to-*stay* as the likelihood that a randomly-assigned enrollee remains in her assigned plan. We also examine which plans auto-assignees switch into, once they make such a switch.

2.3 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 time frame 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 via a round robin approach:[12] Each month, a person in the New York State Department of Health would start from a roster of Medicaid enrollees needing auto-assignment. They would then make assignments to plans in groups of about 20 beneficiaries, using an assignment "wheel." Each group would be assigned to the qualifying plan appearing next on the wheel; then the wheel would cycle until all enrollees were assigned. In a typical month, more than 1,000 enrollees would be assigned in this manner. The following month, assignment would begin again from wherever the wheel had stopped in the prior month.

[12] The sample size of auto-assignees is not identical across plans for several reasons. First, 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. For these reasons, in all specifications we include month $\times$ county of enrollment fixed effects, as within a month $\times$ county of enrollment assignment is purely random.

To be clear, this was not a randomized control trial, and we had no involvement in the randomization process. The quasi-random assignment to plans is a standard part of NY Medicaid administration. We leverage the fact that this policy causes plan choice to be orthogonal to individual characteristics for the subset of the population subject to auto-assignment. Because beneficiaries can opt out of their assigned plans and switch to a different plan, we use an IV research design to address noncompliance. We use assignment to a plan as an instrument for enrollment in that plan. As we show below in Section 3, auto-assignment is a powerful instrument for enrollment, and balance tests—in which data on pre-assignment healthcare utilization allows us to explore correlation between assignment and predetermined characteristics—show no evidence against the assumption that assignment was as good as random.

The limited non-compliance that does occur is driven by the fact that after auto-assignment each beneficiary had three months to switch plans without cause before a nine-month lock-in period began.[13] This is the primary explanation for imperfect compliance, which generates a first stage effect of assignment on enrollment smaller than 1.0, but poses no problem for the maintained exogeneity assumption. Additional institutional details regarding auto-assignment are available in Appendix A and are documented in Wallace (2020), 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.[14] 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

[13] After three months, during the lock-in period, auto-assignees could still switch plans for "cause." Neither form of non-compliance poses any conceptual problem for our IV strategy. As always, compliers (around 90% of beneficiaries here) identify the LATE.

[14] 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.

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 Section 4.2. The expanded-sample results are nearly identical.

These sample restrictions leave us with 65,591 auto-assigned beneficiaries in five boroughs and ten plans. The final "auto-assignee" sample includes 258 month $\times$ county of enrollment (the unit of randomization) cohorts of observations. Table 1 presents summary statistics on this sample.

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 component (ξ_i), time-varying observables (X_{it}), and a mean zero shock.[15] To recover plan effects, γ_j, we estimate regressions of the form:

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

In these regressions, an observation is a beneficiary-month.[16] The regressors of interest are indicators for enrollment in month t in each of the nine plans competing in the New York City market (with the tenth 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 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

[15] 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. The assumed data generating process can be written as $Y_{ijct} = \nu X_{it} + \gamma_j + \xi_i + \epsilon_{ijct}$. The regression combines the individual component, ξ_i, and the error, ϵ_{ijct}, into a compound error term μ_{ijct}.

[16] In Appendix C.2 we present results from regressions where we aggregate to the person-by-six month period instead of the person-month period. Results are similar.

are ten plans that receive auto-assigned enrollees during our time period, requiring nine first-stage regressions (with plan 10 omitted):

$$\begin{aligned} \text{Plan\_1}_{\,ict} &= \alpha_1 + \phi_{1ct} + \delta_1 X_{ict} + \sum_{j=1}^{9} \lambda_{1j} \mathbf{1}[\text{Assigned\_j}_{\,ict}] + \eta_{1,ict} \\ &\vdots \\ \text{Plan\_9}_{\,ict} &= \alpha_9 + \phi_{9ct} + \delta_9 X_{ict} + \sum_{j=1}^{9} \lambda_{9j} \mathbf{1}[\text{Assigned\_j}_{\,ict}] + \eta_{9,ict}. \end{aligned} \tag{2}$$

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 the observation month, relative to the omitted plan. For each first-stage regression, a λ_{kj} equal to one when $k = j$ and equal to zero when $k \neq j$ would indicate perfect compliance. The second stage estimating equation uses the vector of predicted enrollment values ($\widehat{\textbf{Plan}_{ict}}$) from the first-stage regressions:

$$Y_{ijct} = \rho + \psi_{ct} + \nu X_{ict} + \sum_{j=1}^{9} \gamma_j \mathbf{1}[\widehat{\text{Plan\_j}}_{\,ict}] + \mu_{ijct}. \tag{3}$$

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

For some analyses, it is useful to reduce the dimensionality of the problem by grouping together plans. The grouping aids with statistical power, as well as with tractability of certain comparisons. In this modified IV regression specification, 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. These estimating equations take the form:

$$Y_{ict} = \rho + \psi_{ct} + \nu X_{ict} + \gamma_{\text{Low}} \mathbf{1}[\widehat{\text{Low Plan}}_{ict}] + \gamma_{\text{High}} \mathbf{1}[\widehat{\text{High Plan}}_{ict}] + \mu_{ict} \tag{4}$$

where we have divided plans into three groups: low, medium, high, with medium being the omitted category. (We define the groupings below.) The corresponding first stage regressions are analogous to Equation 2.[17]

[17]Specifically, the first stage for enrollment in the "low" group is $\text{Low Plan}_{ict} = \alpha_{\text{low}} + \phi_{\text{low},ct} + \delta_{\text{low}} X_{ict} + \sum_{j \in \text{low, high}} \lambda_{\text{low},j} \mathbf{1}[\text{Assigned\_low}_{\,ict}] + \eta_{\text{low},ict}$.

3.2 First-Stage and Instrument Validity

Figure 2B 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. For example, the estimate of λ_{AA} is 0.924, indicating that the probability of Plan A auto-assignees being observed in Plan A in each of the following six months is 0.924. Across all plans, beneficiaries spend more than ninety percent of beneficiary-months on average in the follow-up period in their assigned plan. The high rate of compliance implies that the local-average treatment effects recovered by IV are unlikely to differ much from average treatment effects for the full auto-assignee sample. Table A1 lists all of the first-stage coefficient estimates, λ_{kj}. The overall first-stage F-statistic is reported in Table 2 and exceeds 7,000.

The statutory goal of the state Medicaid administrator was to randomly assign auto-assignees across the eligible plans. Figure 2A presents a series of randomization tests to assess the IV independence assumption to the extent possible, using information on predetermined characteristics like demographics, as well as pre-randomization medical expenditure. To test for correlations between assignment and predetermined 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 subsample of active-choosers of equal size to the auto-assignee sample to equalize statistical power across the two groups.

The two panels of Figure 2A offer different visualizations of the same underlying balance test regressions. In the left panel, we plot the plan coefficients. Results from the active-chooser regressions are plotted as hollow circles and coefficients from the auto-assignee regressions are plotted as solid circles. To create a comparable scale across dependent variables, all coefficients here are normalized by the standard deviation of the combined set of demeaned plan effects. Importantly, within an outcome (row), a uniform normalization is applied to both the active chooser and auto-assignee samples, so that the spreads of plan effects can be compared. The larger spread apparent among the active chooser plan "effects" indicates that there is strong sorting to plans along predetermined enrollee characteristics among this group.

In the right panel of Figure 2A, we plot for each dependent variable the p-value from an F-test of whether the plan "effects" on predetermined characteristics are jointly different from zero, again separately for the active-chooser and auto-assignee samples.[18] Successful random assignment would

[18]Tabular versions of these results are in Table A2.

tend to generate large p-values, indicating no significant relationship, so large 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 predetermined characteristics. p-values exceed 0.05 for all but one characteristic. The test here is unusually strong: The panel nature of the data and the pre-assignment period during which we observe all healthcare utilization for all beneficiaries in the same fee-for-service program allows us to check for balance on exactly the type of healthcare utilization variables we examine as outcomes below—as opposed to merely a few demographic variables.

The analogous balance estimates for the active-choosers show that plan coefficients on predetermined characteristics are large, and each characteristic is predicted by plan choice with $p < 0.05$. The imbalance among a same-sized random subsample of 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.

The exclusion-restriction in our setting requires that the 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. A more relevant 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. (In contrast, exiting the plan of assignment or exiting the managed care program to enroll in FFS Medicaid would pose no problem as enrollees in these scenarios would remain in our data). We rule out the possibility of differential attrition from the sample directly in the data, showing no evidence of it over our study window (see Figure 6, discussed below).[19]

[19] Tabular versions of plan-level attrition results present a similar story (Table A20).

3.3 External Validity

Our primary causal estimates rely on a sample of Medicaid enrollees that were auto-assigned to Medicaid managed care plans in New York. These enrollees are not a random sample of the Medicaid population, as auto-assignment only occurs if enrollees don't select a plan. It is therefore useful to understand whether auto-assignees differ on observables from active choosers. For example, if auto-assignees are healthier or less-engaged with the healthcare system, our IV estimates would be "local" to consumer types who use relatively little care.

Table A3 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 among individuals using typical levels of care. To maximize generalizability of our estimates, in Section 4 we also include a set of analyses where we re-estimate our primary specification after reweighting the auto-assignee sample to match the full Medicaid population on a rich set of demographic and baseline utilization characteristics, including baseline healthcare spending in the initial months of enrollment while *all* individuals (both active-choosers and auto-assignees) were in the same FFS program.

Beyond generalizability of our estimates from the auto-assignees to the full *New York* Medicaid population, our estimates may also speak to broader MCO heterogeneity in other state Medicaid programs. As noted above, New York is typical in contracting with a variety of plan types, including national for-profits, local for-profits, and local non-profits. If the heterogeneity in causal plan effects we estimate in New York is at least partially tied to plan attributes like for-profit status, our findings may be informative of the broader Medicaid MCO landscape.

4 Plan Effects

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 3. Panel (a) of Figure 3 reports the main result—plan effects on monthly log spending 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 13.1% 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 33.4%, 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. Interestingly, the three lowest-spending plans (B, H, I) are the three for-profit plans in our setting. 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 any utilization each month.

These patterns are robust to alternative specifications and constructions of the dependent variable. Table 2 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 Winsorized spending levels (Table A4), and when we aggregate spending over the entire six-month enrollment spell, rather than analyzing monthly outcomes (Table A5). Below, we find it useful for some analyses to group plans into low- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high-spending plans (Plan D) following Equation 4. The group-level spending differences between low and medium and between medium and high plans are 16 and 11 percent, respectively (Table A6).

The range of these estimates is large.[20] For example, the range of our ten plan effects corresponds to 2.5 times the size of the spending difference between plans with no deductible versus a high deductible (Brot-Goldberg et al., 2017). Yet, our estimates are considerably smaller than the observational, cross-sectional differences in plan spending. To better understand this relationship, Panel (b) of Figure 3 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

[20]External reporting supports large differentials across plans in spending: The Office of the Inspector General examined New York Medicaid managed care plans in 2012 (toward the end of our study period) and found almost a 30 percentage point span (68% to 95%) in medical loss ratios across plans (OIG, 2015). Though not directly numerically comparable to our 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. The underlying cost data for the OIG report (plan financial reports) differ from the claims data we use and so provide independent corroboration.

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 2 and A7.

Figure 3 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 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. We find this even when adjusting for an unusually rich set of observables that include prior healthcare spending in a common FFS plan (which would be typically unavailable as a risk-adjuster for MCO plan effects).

4.2 Disruption and Effects Over Time

How do these effects unfold over time? Are they merely temporary disruptions, or do they reflect long-run effects on healthcare utilization? In Figure 4 we plot month-by-month event study versions of our IV regressions, in which time is relative to the month of auto-assignment. Rather than attempting to estimate nine plan effects interacted with indicators for each month of event time, we group plans together, dividing the ten plans into low-, medium-, and high-spending groups based on the IV spending effects as described above. This both improves statistical power and allows for a simpler visual summary of the time patterns of effects. The specification follows Equation 4, but is estimated separately for each point in relative event time—for each of two months prior to random assignment and for each of six months post assignment.[21] 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.

Figure 4 plots the IV estimates. Some disruption is likely to occur in any new plan transition. It is worth noting, however, that *everyone* in our sample experienced a plan transition from the FFS system to a private MMC plan between month -1 and month 0, so any fixed transition effect would

[21] In particular, $\log(\text{Spending} + 1)^{\tau}_{ict} = \alpha^{\tau} + \phi^{\tau}_{ct} + \delta^{\tau} X_{ict} + \lambda^{\tau}_{\text{low}} \mathbf{1}[\widehat{\text{Low Plan}}_{ict}] + \lambda^{\tau}_{\text{high}} \mathbf{1}[\widehat{\text{High Plan}}_{ict}] + \epsilon^{\tau}_{ict}$, which is estimated separately for each $\tau \in -2, -1, 0, 1, 2, 3, 4, 5$. The groups are low- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high-spending plans (Plan D).

be differenced out in the low versus medium plan comparison. That said, a *differential* disruption between low and medium spending plans could explain our results. If disruption is larger in low spending plans *relative* to medium spending plans (the omitted category), one might expect a greater dip in spending in the earliest post-assignment months that rebounds to zero over time as the low-spending plans converge to the spending levels of higher-spending plans. Panels (a) and (b) use our original sample, with Panel (a) showing only the first 6 months post-assignment and Panel (b) extending up to 12 months post-assignment. As discussed in Section 2, our main auto-assignee sample restricts to observations in the first six months following plan assignment. This is due to the fact that few auto-assignees remain enrolled after the sixth month post-assignment.[22] Thus, there is no change in the composition of auto-assignees over time in Panel (a). Panel (b), on the other hand, allows us to examine longer run impacts but also introduces the possibility of composition bias as the sample becomes unbalanced starting in month 6. For Panels (c) and (d) we generate new smaller 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 the baseline sample and specification (Panel a of Figure 4), effects do appear somewhat larger in the first months post-assignment, but they are still large and significant by month 5. Further, the (insignificant) suggestion of attenuation over time in Panel (a) is not replicated across alternative specifications, including in Panel (b) which uses the same sample but allows the horizon to run out an additional six months. In Table A8 we present regression estimates for these different samples, pooling over all post-assignment months. The results in this table show that our estimates of the causal effects of low-spending plans on spending are remarkably consistent across these samples. Overall, Figure 4 and Table A8 show that spending effects remain large throughout the post-assignment months. The implication is not that disruption effects are unimportant, but rather that any time pattern of disruption is common across plans and that differential disruption from being assigned to a low-spending plan is unlikely to explain our results.

[22]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, the modal beneficiary remains enrolled in Medicaid for only 6 months post-assignment, with over 30% of auto-assignees being enrolled for exactly 6 months. Only a few auto-assignees remain enrolled past 6 months after assignment.

4.3 Heterogeneity

In Figure A2 we plot coefficients of plan effects estimated separately in various subsamples of the auto-assignee sample. The three panels split the data by sex, median age, and baseline spending, where the latter is measured prior to assignment to a managed care plan, when enrollees received all care through the FFS system. Differences in plan effects by sex are mostly negligible. Differences by age and baseline spending are more substantial, with larger plan effects estimated for older and sicker groups. The regressions are underpowered to detect statistical differences across plans-by-groups, but the point estimates suggest that the overall plan effects are larger for sicker beneficiaries, proxied here by those who have used more care in the past, but still meaningful for healthier groups.

The final panel in Figure A2 aggregates the data differently, in order to gain statistical power and reveal the time pattern of effects. Here, as in Figure 4, the specification follows Equation 4, grouping plans into high-, medium-, and low-spending groups and allowing for heterogeneous treatment effects over event time. The figure plots the low (versus medium) coefficients estimated separately for two subsamples: those with no spending in the baseline period and those with positive spending. Consistent with the plan-level estimates, the impact of being assigned to a plan in the low-spending group (relative to the medium group) is largest for the sicker beneficiaries. The differences are statistically significant at the beginning of the event window and marginally significant at the end. Appendix Table A9 presents pooled regression results corresponding to each of the event studies in Figure A2, revealing that at this level of aggregation there is little heterogeneity in plan effects by age and sex but significant heterogeneity by baseline spending: Spending effects are 60% larger for beneficiaries with some baseline spending relative to beneficiaries with no baseline spending. Clearly, spending effects are not driven by healthy beneficiaries with minimal interaction with the healthcare system. Instead, effects are driven by sicker beneficiaries who frequently use care.

5 Prices, Quality, and Satisfaction

In this section we evaluate whether the relative savings of lower-spending plans were associated with observable correlates of clinical plan quality and/or revealed enrollee preference. Because lower negotiated provider prices could cause spending reductions without implying a reduction in plan quality or consumer welfare, we begin by decomposing the role of prices.

5.1 Prices

In New York's MMC program, different plans have the ability to negotiate prices (for identical services) that differ from each other and from the fee-for-service program. The extent to which these prices actually differ, however, is an empirical question. The similarity of plan effects on total spending and plan effects on an indicator for any utilization (Figure 3) suggests that quantity differences may be more important than negotiated price differences in explaining spending differences in this context. Figure 5 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.[23] For ease of comparison, Panel (c) of Figure 5 plots these plan effects re-estimated on the price-standardized data against our main IV plan effects. (These coefficients are tabulated in Table A4.) 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 first panel of Figure 6, we summarize pricing results for plan groupings (low-, medium-, and high-spending). Each row of the figure corresponds to a separate regression, plotting coefficients and 95% confidence intervals for the group-level coefficients on low-spending plans. The first coefficient is from the unadjusted regression (also in Table A6). The next row reports the coefficient on price-standardized spending that parallels panel (c) of Figure 5, and shows no change. The final row in the first panel reports the coefficient from a price-standardized regression in which we additionally reprice each denied (zero paid) claim as if it had been paid at a common-across-plans price, and then re-estimate the effect of low- versus medium-spending plans. Denied claims also do not appear to be important for explaining spending effect differences. These results combine to show

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

that low-spending plans are reducing spending by reducing actual utilization of healthcare goods and services, *not* by just paying less for the same goods and services.

5.2 Marginal Services

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. In our setting, are the utilization reductions achieved using non-cost-sharing tools similarly broad-based, or are the services that are marginal to enrollment in lower-spending plans more targeted—and perhaps of lower value?

In the remaining panels of Figure 6, we investigate whether the reductions in spending generated by managed care are similarly blunt or better targeted. We begin in the service panel by examining plan effects by type of service. Each row reports an IV coefficient estimate on low-spending plan enrollment. The dependent variables in the panel are indicators for any use of the service type in the enrollee-month, and coefficients are divided by the mean of the dependent variable in the omitted group in order to place multiple outcomes on the same scale. The panel 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. Beneficiaries assigned to the low-spending plans also used fewer emergency department (ED) visits, consistent with evidence that for some populations ED may be a complement to, rather than substitute for, other ambulatory care (Finkelstein et al., 2012; Cuddy and Currie, 2020).

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 (e.g., ED utilization). To investigate this, we examine two sets of potentially high-value services that could produce spending offsets: high-value drugs and high-value services, including primary care.

The drug and high-value care panels in Figure 6 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.

The high-value care panel of Figure 6 analyzes six measures of compliance with recommended care developed by the Department of Health and Human Services for Medicaid enrollees: the use of primary care, the prevalence of HbA1c testing, breast cancer screening rates, cervical cancer screening rates, chlamydia screening rates, and flu vaccination rates. For primary care and breast cancer screening, there is no difference in spending. The coefficient for primary care, in particular, is a precisely estimated zero. For flu vaccinations, the effect is negative but insignificant. Among the other measures, we find that enrollment in a low-spending spending plan significantly reduces the use of recommended preventive care. In sum, there is no indication that low-spending plans achieve savings by promoting high-value care and achieving offsets. Instead, similar to what happens when consumers face a high deductible, supply-side managed care tools appear to constrain most of care with the exception of primary care (Figure A4), which we return to in Section 5.4.

Beyond plan effects on high-value services, we also 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, chest imaging, and head imaging for an uncomplicated headache (Schwartz et al., 2014; Charlesworth et al., 2016). With the exception of possibly reducing overall imaging (but not narrowly defined low-value imaging), the low-value care panel of Figure 6 shows no evidence that low-spending plans reduce the use of these low-value services. These results are somewhat in contrast to the finding that lower-spending plans make across-the-board reductions by service setting (inpatient, clinic, pharmacy, etc.), but they make it very clear that these plans are not selectively cutting out services that offer little value to patients. Indeed, these are the few services where utilization appears *not* to be affected by low-spending plans, though we note that confidence intervals are wide, leaving us unable to rule out significant decreases as well as significant increases in the use of these services.

Finally, as another dimension of heterogeneity, we can examine differences across plans in enrollee spending on services carved out of MMC plan contracts and always paid by the FFS program, even for beneficiaries enrolled in MMC. A minority of services for managed care enrollees are carved-out and paid directly by the state on a fee-for-service basis. The claims data for these services are generated by the state and merged with the plan data. If carved-out services are substitutes for carved-in services, low-spending plans may strategically push beneficiaries to use carved-out services (for which plans bear no financial responsibility) in place of carved-in services (for which plans are the residual claimant). On the other hand, if plans impact spending on both carved-in and carved out services similarly then plan effects may show up even in carved-out FFS claims. In Appendix Figure A5, we estimate the IV plan effects on each spending component separately. The figure shows that the patterns of plan effects on FFS claims are tightly correlated with the patterns of managed care claims. Either there are important complementarities between managed-care-paid and FFS services, or cost-saving reductions are blunt, rather than strategically targeted.

Importantly, these results also carry the implication that plan spending differences are unlikely to be driven by differential reporting. The FFS services represent a data component that cannot be contaminated by plan reporting differentials. The plans themselves have no reporting role for these claims, yet we observe a tight correlation between plan effects on self-reported spending (via managed care encounter data) and plan effects on state-reported spending (via FFS claims), providing strong evidence that the differences in self-reported spending are not merely due to differential reporting.

5.3 Satisfaction and Health

In the Medicaid setting, beneficiaries enrolling in lower-spending plans are not subject to cost-sharing. Hence, the classic trade-off between financial risk protection and moral hazard (Zeckhauser, 1970) is absent. There may, however, be a trade-off between satisfaction and plan spending, as well as a potential trade-off between spending and health. We study the spending/satisfaction 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 after auto-assignment rather than switch to a different plan. Recall that enrollees can switch away from their plan of assignment. In the language of IV, these are never-takers with respect to the auto-assignment instrument. Random assignment allows us to interpret empirical differences in the likelihood of switching plans as causal effects of being assigned to those plans.

We operationalize this measure of beneficiary satisfaction as the probability that an auto-assignee remains enrolled in their assigned plan and call it “willingness-to-stay." The key assumption underlying the interpretation is the typical one: that choices (to remain enrolled or switch plans) reveal preferences. Unlike willingness-to-pay, our measure of beneficiary satisfaction is not scaled to dollars, but rather reports probabilities of continued enrollment. Despite the scale limitations, one potential advantage of our measure is that it plausibly offers some insight on consumers' *experienced* utility in a plan, as it is measured as a reaction to (i.e., causal effect of) being enrolled in a plan. For certain questions related to ex-post consumer evaluations, willingness-to-stay may be preferable to, for example, a willingness-to-pay measure derived from initial plan choices in a market setting with important information frictions.

The enrollee satisfaction panel of Figure 6 shows that people are less likely to stay in lower-spending plans. Willingness-to-stay is lower in these plans and declines over the post-assignment window, reaching a differential of several percentage points (relative to willingness-to-stay in medium-spending plans) by six months post-assignment. This is consistent with enrollees learning about the poor subjective quality of low-spending plans over time. Appendix Figure A6 shows the analog for the one high spending plan. There, as well, beneficiaries are similarly more likely to stay in the high-spending plan versus the medium-spending plans. It's not possible to directly observe in these claims data whether the revealed dissatisfaction reflects difficulty scheduling appointments, restrictive gate-keeping by PCPs, or other factors—though we discuss the possible roles of these and other factors in Section 5.4.

To give a finer view of these results, in Figure 7 we plot plan-level estimates of willingness-to-stay against the plan effects on spending. The relationship is clear, with higher-spending plans having higher estimates of willingness-to-stay. In Appendix Figure A7 we present a similar figure, stratifying the auto-assignees by whether or not that had any baseline spending. This figure suggests that sicker beneficiaries—those who use more care and so have more experience with their plans—drive the relationship. Thus, the plan effects we estimate via the claims data are strongly correlated with consumers' actual experiences in the plans and their decisions over continued enrollment, consistent with a binding trade-off between plan spending and beneficiary satisfaction.

To investigate the trade-off between spending and health, we use a standard, if imperfect, 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 6 shows that enrollees in the low-spending plans are 15% *more* likely to have an avoidable hospitalization despite having lower utilization for most other types of care. This result is particularly striking in the context of our prior results showing that for the vast majority of healthcare services, low-spending plans generate lower levels of utilization. This result shows that in contrast to most healthcare services, when it comes to types of services whose utilization may indicate a deterioration of beneficiary health, low-spending plans generate *higher* levels of utilization. This suggests that the tools used by low-spending plans to constrain costs could have negative consequences for beneficiary health.

5.4 Summary and Potential Mechanisms

To summarize, our results show that even without exposing consumers to out-of-pocket spending, plans exert significant influence over total spending. In this sense, supply side interventions by plans—as opposed to consumer cost-sharing—can constrain healthcare spending while circumventing the classic trade-off between financial risk protection and moral hazard (Zeckhauser, 1970; Pauly, 1974). However, those reductions are not a free lunch, with costs borne by beneficiaries in terms of the quality of care delivered, health outcomes, and in a revealed preference measure of satisfaction. Further, a key limitation of reducing spending via consumer cost sharing is replicated here: The impacts are blunt and broad-based, rather than targeted to low-value services.

Next, we briefly explore what we can learn about *how* plans achieve spending reductions. Since there is no consumer cost sharing in our setting, and the statutory scope of covered benefits is set by the state, causal differences in spending between plans must be driven by differences in their use of supply-side (i.e., managed care) tools. Though the term *managed care* can encompass a wide range of mechanisms, Glied (2000) summarizes the key methods as the selection and organization of providers (i.e. networks), how plans negotiate payments to providers, and utilization management, in its various forms. We thus take a moment to briefly explore the potential channels through which the lower-spending plans constrain costs, to the extent possible given our data and setting, and discuss the implications of our findings for the economics of Medicaid managed care.

We start with the caveat that no research design, including ours, is likely to be well-suited to simultaneously estimating general equilibrium plan effects and to isolating which mechanisms (holding all else fixed) are most important for explaining those plan effects. Random assignment of enrollees to plans identifies the impacts of *plans*, rather than specific plan features. That is because plans are bundles of mechanisms, generated endogenously by market and regulatory processes. An advantage of our identification strategy is that we can estimate the size of causal plan effects for plans as they are—in equilibrium and at scale. A disadvantage is that randomization of people to plans cannot isolate the efficacy of any single plan feature. Practically speaking, even if one made the (poor) assumption that bundles of plan features were as good as randomly distributed across plans as these arise "in the wild," dimensionality would impose a binding constraint on projecting plan effects onto plan features because plans differ on more dimensions than the number of plans competing in any market. Even our exceptionally dense MMC market setting includes only ten plans/carriers, which may differ in dozens of important ways. With this caveat in mind, we review the potential mechanisms and available evidence in our setting, following Glied (2000).

One way that plans can constrain healthcare spending is through the selection and organization of contracted providers. Medicaid managed care plans are given substantial leeway to construct their networks of contracted providers (e.g., physicians and hospitals). Networks, which do vary significantly across the 10 plans we study (Wallace, 2020), can statistically explain *some* of the plan heterogeneity in spending we observe, but the correspondence isn't particularly strong (see Appendix Figure A8).[24] In fact, though the relationship is positive overall (broader networks are associated with higher spending) two of the highest spending plans in New York City Medicaid included one narrow network, vertically-integrated plan and one wide network plan, complicating any simple attribution of plan effects to network breadth. The relationship between our causal plan effects and network breadth strengthens if we control for vertical integration by removing the provider-owned plan, but the narrower network plans also tended to be the only for-profit plans in the market (Table 2), highlighting the difficulty of disentangling mechanisms in our setting.[25]

A more subtle view of the importance of networks, beyond the question of broad versus narrow, involves whether certain plans are better at steering patients to providers with more efficient practice

[24]Appendix Tables A10-A13 demonstrate that our causal estimates of plan effects are also qualitatively similar if we control for provider network breadth by using plan of assignment interacted with enrollee zip code to instrument for network breadth (which varies at the plan-by-zip level). See Appendix Section D.1 for additional details.

[25]See Appendix Section D.1 for full detail on the network analysis. We also investigate there the relationship between our plan effects and another dimension of provider network breadth, differences in how *binding* the network restrictions are in each plan. To do this, we examine the correlation between plans' causal effects and a plan-specific, out-of-network (OON) hassle cost estimated from the hospital demand model in Wallace (2020). Plans differed in how difficult it was to access OON hospital care, and we find suggestive evidence of a linear relationship between OON hassle costs and willingness-to-stay (with enrollees less satisfied in more restrictive networks). See Panel (b) of Appendix Figure A8.

styles (Glied, 2000). To assess whether efficient steering could explain our results, we attribute enrollees to providers (based on where enrollees utilize care) and re-estimate our primary specification with provider fixed effects (see Appendix Section D.1 for details). The analysis shows that provider fixed effects, too, can explain little (only 10-20%) of the variation in plan spending, suggesting that lower-spending plans aren't achieving savings primarily by differentially selecting certain providers to contract with (Appendix Tables A10-A11). Further, plans are not paying providers differently in a way that could account for our findings; Figure 5 showed plan differences in negotiated provider rates were not strongly correlated with plan spending. That finding, combined with our findings that real resource utilization varied across plans, rules out the possibility that similar care was simply reimbursed differently across plans.

Another candidate explanation that has been highlighted by the literature, but for which we find little evidence, is case management, such as AI-targeted follow-up, coordinating referrals to a specialist, post-discharge planning, and proactively routing patients to high-value care. These are meant to reduce costs by improving health and reducing adverse events. In our context, we see no evidence of increased use of high-value services in lower-spending plans and worrying evidence of deteriorating health in the form of higher avoidable hospitalization rates.

An important residual that we cannot directly observe is utilization management—e.g., prior authorization for certain kinds of care and simply restricting access to services and technologies. Restricting access would be consistent with our findings of large effects on the extensive margin of any use in an enrollee-month,[26] but claims and encounter data, such as we use here and are typically used in econometric studies of health insurance markets, can offer no direct evidence on this issue. (An ideal dataset would document interactions *outside of the insurance claim workflow*.) Because the circumstantial evidence we provide above rules out many alternative explanations—prices, denials, case management, networks, and provider steering—our findings speak to the importance of developing new datasets for research aimed specifically at understanding utilization management.

[26]This explanation would also be consistent with the lack of evidence that lower-spending plans reduce the likelihood of using primary care (e.g., see Figure A4). This is because supply-side managed care tools (e.g., prior authorization, step therapy, etc.) are generally designed to restrict access to downstream care (e.g., outpatient specialty care, imaging and lab, etc.), conditional on being seen by a primary care provider.

6 Consumer Demand and Implications for the Economics of Medicaid Managed Care

In this section, we examine the relationship between Medicaid managed care plan performance and plan market shares. An important feature of Medicaid is that, unlike in other health insurance markets, beneficiaries pay no premiums and face no cost-sharing for care. In the absence of prices, enrollees may choose plans based on other attributes (e.g., clinical quality, access to care, or customer service). However, there are many reasons to doubt that enrollment flows necessarily follow clinical or other measures of plan quality, such as publicly-reported plan report cards, given the type of choice frictions and imperfections often documented in this domain (e.g., Handel and Kolstad, 2015; Abaluck et al., 2021). In this section, we show that consumer demand drives greater market share to the MMC plans that generate higher causal plan spending (as estimated in Figure 3). Publicly-reported plan ratings, on the other hand, do not correlate with plan market shares. We discuss the implications of these allocation results for the roles of managed competition and procurement in Medicaid Managed Care.

6.1 Allocation Results

Figure 8 presents our central result on the static allocation of enrollees across plans. Panel (a) shows a strong correspondence between our causal spending effects (on the x-axis) and mean plan market shares among the broader population of active choosers during our study period (on the y-axis). Consumer demand follows spending: Health plans with 10% higher spending among the randomly-assigned enrollees have a 4.1 percentage point (41 percent) higher market share among enrollees making active choices (R^2 of 0.77). The greater allocation of enrollees to higher-spending plans is consistent with evidence that consumer demand follows quality in the hospital sector (Chandra et al., 2016), as these plans do less to restrict access to services and also seem to improve clinical quality. However, panel (b), which plots the same set of market shares against the regulator-reported overall plan rating, shows that enrollee choices *do not* seem to follow publicly-reported plan quality (R^2 < 0.01). These ratings are the one piece of information about plans provided to beneficiaries by Medicaid at the time of choice. Taken together, the Figure 8 results show that demand seems to primarily be tied to the ability to use care, and thus to higher levels of healthcare spending.

Figure 9 illustrates the different channels by which consumer demand follows plan spending by plotting the flows of both active chooser and auto-assignee beneficiaries into and out of plans. Plans in each panel are ordered from left-to-right in terms of highest to lowest estimates of causal effects on spending. Panel (a) shows plan retention of auto-assignees across these plans, based on the 6-month willingness-to-stay coefficient estimates already presented in Table 2. Panel (b) presents a new result: the destination plans among auto-assignees who switch plans after initial assignment. Panel (c) shows the initial plan choices of all active-choosers (those who are not auto-assigned). And panel (d) parallels (b) for active choosers, showing the destination plans among active-choosers that switch plans after their initial selection.

The four measures of plan preference in Figure 9 tell a remarkably consistent story across all four pathways: The highest spending plans tend to be the most preferred. That is true across the cases of active choosers, auto-assignees, initial plan choices, plan retention, and the destination plans among plan switchers. Three of the four highest spending plans dominate enrollment flows. (The fourth, Plan E, is dis-preferred similarly among all groups, suggesting it represents an interior point on a production possibilities frontier defined by costs and consumer preferences.) Lower-spending plans cannot offer enrollees lower prices to compensate them for reduced service provision, with the predictable result that low-spending plans attract smaller enrollment shares—at least among the attentive beneficiaries generating variation in Figure 9.[27] The lowest spending plan, I, attracts the lowest (nearly zero) enrollment shares by any of the flow measures in Figure 9. In contrast, among auto-assignees, plans receive equal *initial* shares of enrollment flows by the randomization design.

6.2 Competition and Consumer Choice May Drive Program Costs (Up)

An important implication of our allocation results is that consumer choice in this context drives higher Medicaid program costs, offering a potential resolution to the puzzle of rising program costs under privatization. To illustrate, consider a counterfactual that reallocated all active-chooser beneficiaries randomly across plans instead of according to the actual, endogenous market shares observed

[27] One might ask why some plans seem to pursue different strategies than others in this market. It is conceivable that the low-spending, low-enrollment plans are pursuing a high-margin low-volume strategy that contrasts to the low-margin, high-volume market niche of the high-spending plans. Such an equilibrium is possible under a model where some beneficiaries actively choose plans that match their preferences while other beneficiaries either choose somewhat randomly (because of very limited information at the time of plan choice) or who are literally randomly assigned to plans because they neglect to make an active choice. Many states have high auto-assignment/passive choice rates, making a large portion of choices in the market literally random and completely insensitive to plan design.

in the data. Holding all else fixed, this would reduce healthcare spending by about $559 million annually in New York City Medicaid alone (4.6% or $254 per active-choosing beneficiary per year), per our main plan spending effects in Figure 3, combined with enrollment shares.[28] That is a substantial fraction of program costs, though the way reduced plan spending passes-through into program costs is indirect: Regulators must set actuarially-sound capitation rates, which requires certification by actuaries that forecast future spending based on current (and past) plan spending; hence, if demand allocates more enrollees to higher-spending plans today, that leads to states paying higher plan capitation rates tomorrow (Spitz, 2007).[29]

States have policy tools at their disposal to counteract the tendency for managed competition to drive up program costs. First, states could redesign their auto-assignment algorithms with the goal of improving the efficiency of the Medicaid program (Wallace, 2020).[30] In many state programs, auto-assignees are a large fraction of the total Medicaid population. For example, of the 37 MMC states for which Kaiser Family Foundation (2015) collected data on auto-assignment rates, the median statewide rate of auto-assignment is over 40% of the Medicaid Managed Care population. Given that beneficiary compliance with auto-assignment is high, algorithms that favored lower-spending plans could reallocate enrollees to lower-spending plans to incentivize greater cost control (Marton, Yelowitz and Talbert, 2017; **?**). Second, states could afford MMC plans greater flexibility to offer additional services to beneficiaries via two existing channels: "in lieu of services" or "value-added services."[31] This would allow lower-spending plans to attract greater market share by reinvesting savings in services beneficiaries value.[32]

[28]The active chooser-enrollment-weighted average monthly spending effect is -0.031. The unweighted average monthly spending effect is -00.077. We calculate the dollar value of per-beneficiary-year savings from random reallocation as $((-0.031) - (-0.077)) * 5532 = 254$, where $5,532 is the average annual spending among active-choosers. We multiply this by 2.2 million active choosers.

[29]This link between plan spending and state spending is highly salient to state policymakers. Based on our conversations with state Medicaid administrators, even in settings like ours where plan payments are capitated, there is a clear focus on keeping managed care plan spending down for this explicitly-stated reason.

[30]Today, many states allocate auto-assignees at random (among qualifying plans). In some cases, allocations are equal across the plans, but in other cases they are proportional to market shares (to maintain plan shares) or, alternatively, intentionally targeted to smaller plans (to support new entrants). The clear lack of a best practice suggests that smarter auto-assignment policies may be a substantial missed opportunity.

[31] Under 42 C.F.R. 438.3(e), which provides the authority for the provision of "in lieu of services," a state Medicaid agency may permit its contracted Medicaid managed care plans to cover one or more services which are not covered under the state's Medicaid State Plan if these services are medically appropriate and cost-effective substitutes for State Plan-covered services. For example, the state may authorize plans to cover home visiting services for pregnant women "in lieu of" prenatal visits at a clinic or doctor's office.

Under 42 C.F.R. 438.3(e)(1)(i), which provides the authority for the provision of value-added services, Medicaid managed care plans may offer members additional services that are not substitutes for State Plan services. These may include, for example, subsidies for exercise classes, online mental health resources, or nutrition counseling.

[32]It could also generate a mechanism by which plans could engage in service-level selection, contributing to socially

In sum, these observations about competition for managed care enrollees, combined with our findings, raise concerns about the capacity for managed competition—in its current form in Medicaid—to drive cost savings. This may help explain the lack of savings associated with privatization in Medicaid, and is consistent with recent work finding that expansions of managed care to disabled Medicaid beneficiaries led to higher program costs (Layton et al., 2019; Layton and Politzer, 2021)[33]

6.3 Procurement

The procurement process offers an additional lever states can use to manage costs. Currently, states exert varying levels of control over which managed care plans participate in their Medicaid programs with the entire procurement process differing a great deal across states (see Layton, Ndikumana and Shepard (2017) for a full description). Some states allow most, if not all, interested firms to participate, using their procurement power only occasionally as a way to punish plans that do not meet some minimum quality threshold. Other states are much more restrictive, stating *ex ante* that they will select a limited number of firms (often around three), and all eligible Medicaid beneficiaries will be required to enroll in the plans offered by those firms. Sometimes the procurement involves statewide contracts, while in other cases procurement occurs at the regional level. Contracts tend to last 3-5 years and involve the possibility of renewal.

Our results suggest that the state's choice of which firms to contract with does indeed matter for program costs. To illustrate, consider a counterfactual that is more radical than the counterfactual in Section 6.2: A reallocation that removed the top 4 highest-spending plans in the market and re-assigned their enrollees to the remaining 6 plans according to the existing market shares of those 6 plans. Such a reallocation, holding other features fixed, would reduce healthcare spending by \$1.4 billion annually in New York City Medicaid alone (11.5% or \$637.56 per active-choosing beneficiary per year).[34] Thus, procurement matters for the cost of managed care.

Our results also indicate that states' procurement decisions carry a real trade-off. If states want to maximize quality and satisfaction, they can do so by selecting the higher utilization plans. Doing

inefficient screening (Geruso, Layton and Prinz, 2016).

[33]At the same time, if states place little weight on costs and high weight on quality, access, and beneficiary health and satisfaction, our results suggest that competition is likely to achieve the desired program outcomes.

[34]The active chooser-enrollment weighted average monthly spending effect is -0.031. The active-chooser weighted average monthly spending effect restricting to the bottom 6 low-spending plans is -0.146. We calculate the dollar value of per-beneficiary-year savings from removing the top 4 plans from the market as $((-0.031) - (-0.146)) * 5532 = 637$, where \$5,532 is the average annual spending among active-choosers. We multiply this by 2.2 million active choosers.

so will lead to higher spending. Similarly, if states want to minimize spending, they can do so by selecting the lower-spending plans (disproportionately for-profits in our setting). Doing so will lead to lower satisfaction and access to care. It may also lead to more avoidable hospitalizations and other undesirable outcomes that are more difficult to measure. Figure 7 suggests some scope for using procurement to select more efficient plans (i.e., those generating higher consumer satisfaction at the same level of spending), but limited statistical power prevents us from making firm conclusions here.

Finally, our results provide some guidance regarding how to select plans that best align with state objectives even in settings where it is not possible to estimate causal plan effects as we have done here. Panel (b) of Figure 3 shows that our quasi-experimental estimates of plan spending effects are highly correlated with risk-adjusted OLS estimates, suggesting that the ordering of OLS estimates is informative, even if the level differences are misleading.[35] This suggests that states might be able to achieve spending reductions by simply selecting plans with the largest negative OLS risk-adjusted spending effects.[36]

7 Conclusion

What difference does a plan make? Using large-scale random assignment to the ten plans participating in one of the largest Medicaid managed care markets in the US, we show that a plans can indeed make a difference, both for consumers and for the cost of the public programs serving them. The range of spending and utilization impacts across managed care plans in our setting is around 30 percentage points. The impact on program costs and real resource use of enrolling in a lower-spending plan in place of a higher-spending plan is thus larger than what could be accomplished by exposing consumers to high deductibles and reasonable coinsurance and copays.[37] In this way, managed care circumvents the classic trade-off between financial risk protection and moral hazard noted by Zeckhauser (1970) and Pauly (1974). Our findings are particularly relevant for public insurance programs—including Medicaid, the low-income segments of the HIX Marketplaces, and the

[35] The OLS estimates are generally larger than the quasi-experimental estimates, suggesting that there is adverse selection into the plans that cause higher spending and advantageous selection into plans that cause lower spending.

[36] Further, recall that the higher-spending plans had higher levels of access and satisfaction, as well as higher levels of avoidable hospitalizations, suggesting that that OLS risk-adjusted spending estimates may provide a good signal of plan causal effects on quality, satisfaction, and health, and that states may be able to select plans with the best effects on quality by simply selecting the plans with the highest risk-adjusted spending levels.

[37] Evidence on deductibles and coinsurance has been clearly documented by Manning et al. (1987), Brot-Goldberg et al. (2017), and others.

Low Income Subsidy Program in Medicare Part D—where policymakers have been reluctant to expose low-income consumers to financial risk, or in some cases, reluctant to expose these consumers to cost-sharing in any form.

We also show that, somewhat contrary to popular claims, achieving healthcare savings via managed care offers no free lunch. Consumer satisfaction—as captured in the revealed preference decision to remain enrolled in an assigned plan—is strongly negatively correlated with a plan's cost savings. And reductions caused by lower-spending plans are blunt: Lower-spending plans reduce utilization of all types of care, generating low scores on traditional measures of healthcare quality and increasing the likelihood of adverse health events. There is almost no evidence in our study that supports the idea that managed care substantially reduces costs by steering patients toward higher value care or by keeping patients healthy.

Finally, our findings carry important implications regarding the potential for managed care plans to constrain healthcare spending growth in Medicaid. Medicaid beneficiaries face a choice of managed care plans but do not face different prices for enrolling in different plans. We document a close link between plan spending and beneficiary demand that implies competition is likely to drive up program costs as consumers favor the higher spending options. While there are policy avenues available to counteract this tendency—in particular targeted auto assignment and active procurement—these facts make it difficult for states to reign in costs without limiting choice. As managed care continues to evolve, it will be important for future work to continue to critically evaluate and document whether and how managed care generates real efficiencies in healthcare consumption. Taken as a whole, our results show that plans matter considerably for spending and satisfaction, but are primarily choosing different points along the cost and quality frontier—not pushing it outward.

References

Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. 2021. "Mortality effects and choice across private health insurance plans." *The quarterly journal of economics*, 136(3): 1557–1610.

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.

Angrist, Joshua D, Peter D Hull, Parag A Pathak, and Christopher R Walters. 2017. "Leveraging lotteries for school value-added: Testing and estimation." *The Quarterly Journal of Economics*, 132(2): 871–919.

Angrist, Joshua, Peter Hull, Parag Pathak, and Christopher Walters. 2016. "Interpreting tests of school VAM validity." *American Economic Review*, 106(5): 388–92.

Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein. 2013. "The RAND Health Insurance Experiment, Three Decades Later." *Journal of Economic Perspectives*, 27(1): 197–222.

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.

Byrd, Vivian LH, Allison Hedley Dodd, et al. 2015. "Assessing the usability of encounter data for enrollees in comprehensive managed care 2010-2011." Mathematica Policy Research.

Cabral, Marika, Michael Geruso, and Neale Mahoney. 2018. "Do larger health insurance subsidies benefit patients or producers? Evidence from Medicare Advantage." *American Economic Review*, 108(8): 2048–87.

Centers for Medicaid and Medicare Services. 2011. "Medicaid Managed Care Enrollment Report." *https://www.kff.org/wp-content/uploads/sites/2/2013/12/2011-medicaid-mc-enrollment-report.pdf*.

Centers for Medicaid and Medicare Services and Mathematica Policy Research. 2013. "Medicaid Managed Care Enrollment and Program Characteristics, 2013." *https://www.kff.org/wp-content/uploads/sites/2/2013/12/2011-medicaid-mc-enrollment-report.pdf*.

Centers for Medicaid and Medicare Services and Mathematica Policy Research. 2018. "Medicaid Managed Care Enrollment and Program Characteristics, 2018." *https://www.medicaid.gov/medicaid/managed-care/downloads/2018-medicaid-managed-care-enrollment-report.pdf*.

Centers for Medicare and Medicaid Services. 2016. "Medicaid Managed Care Enrollment and Program Characteristics."

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health care exceptionalism? Performance and allocation in the US health care sector." *American Economic Review*, 106(8): 2110–44.

Chandra, Amitabh, and Douglas O Staiger. 2020. "Identifying sources of inefficiency in healthcare." *The quarterly journal of economics*, 135(2): 785–843.

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.

Chernew, Michael, J Sanford Schwartz, and A Mark Fendrick. 2015. "Reconciling prevention and value in the health care system." *Health Affairs. Blog Post*.

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.

Cuddy, Emily, and Janet Currie. 2020. "Rules vs. discretion: Treatment of mental illness in us adolescents." National Bureau of Economic Research.

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.

Curto, Vilsa, Liran Einav, Jonathan Levin, and Jay Bhattacharya. 2021. "Can health insurance competition work? evidence from medicare advantage." *Journal of Political Economy*, 129(2): 570–606.

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.

Dranove, David, Christopher Ody, and Amanda Starc. 2017. "A Dose of Managed Care: Controlling Drug Spending in Medicaid." National Bureau of Economic Research Working Paper 23956.

Duggan, Mark, Amanda Starc, and Boris Vabson. 2016. "Who benefits when the government pays more? Pass-through in the Medicare Advantage program." *Journal of Public Economics*, 141: 50–67.

Duggan, Mark, and Tamara Hayford. 2013. "Has the Shift to Managed Care Reduced Medicaid Spending? Evidence from State and Local-Level Mandates." *Journal of Policy Analysis and Management*, 32(3): 505–535.

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.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. 2012. "The Oregon health insurance experiment: evidence from the first year." *The Quarterly journal of economics*, 127(3): 1057–1106.

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

Garthwaite, Craig, Christopher Ody, and Amanda Starc. 2020. "Endogenous Quality Investments in the U.S. Hospital Market." National Bureau of Economic Research Working Paper 27440.

Gaynor, Martin, and Robert J Town. 2011. "Competition in health care markets." *Handbook of health economics*, 2: 499–637.

Geruso, Michael, and Timothy Layton. 2017. "Selection in Health Insurance Markets and Its Policy Remedies." *Journal of Economic Perspectives*, 31(4): 23–50.

Geruso, Michael, Timothy J. Layton, and Daniel Prinz. 2016. "Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges." National Bureau of Economic Research Working Paper 22832.

Glied, Sherry. 2000. "Managed care." In *Handbook of health economics*. Vol. 1, 707–753. Elsevier.

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, and Jonathan Kolstad. 2015. "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review*, 105(8): 2449–2500.

Handel, Benjamin, Jonathan Holmes, Jonathan Kolstad, and Kurt Lavetti. 2019. "Insurer Innovation and Health Care Efficiency: Evidence from Utah." Work in Progress.

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

Kaiser Family Foundation. 2001. "Medicaid and Managed Care." https://www.kff.org/wp-content/uploads/2013/01/medicaid-and-managed-care-fact-sheet.pdf.

Kaiser Family Foundation. 2010. "Medicaid and Managed Care: Key Data, Trends, and Issues." https://www.kff.org/wp-content/uploads/2012/02/8046-02.pdf.

Kaiser Family Foundation. 2015. "Medicaid Reforms to Expand Coverage, Control Costs and Improve Care: Results from a 50-State Medicaid Budget Survey for State Fiscal Years 2015 and 2016." https://files.kff.org/attachment/report-medicaid-reforms-to-expand-coverage-control-costs-and-improve-care-results

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, Alice Ndikumana, and Mark Shepard. 2017. "Health Plan Payment in Medicaid Managed Care: A Hybrid Model of Regulated Competition." National Bureau of Economic Research Working Paper 23518.

Layton, Timothy, and Eran Politzer. 2021. "âĂIJThe Fiscal Consequences of Private Provision of Medicaid: Evidence from Enrollment Mandates for Adults with Disabilities." *Mimeo*.

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.

Marton, James, Aaron Yelowitz, and Jeffery C Talbert. 2017. "Medicaid program choice, inertia and adverse selection." *Journal of health economics*, 56: 292–316.

Marton, James, Aaron Yelowitz, and Jeffery Talbert. 2014. "A Tale of Two Cities? The Heterogeneous Impact of Medicaid Managed Care." *Journal of Health Economics*, 36(1): 47–68.

Medicaid, CHIP Payment, and Access Commission. 2020. "MACStats: Medicaid and CHIP Data Book." https://www.macpac.gov/wp-content/uploads/2020/12/MACStats-Medicaid-and-CHIP-Data-Book-December-2020.pdf.

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.

Song, Zirui, Mary Beth Landrum, and Michael E Chernew. 2013. "Competitive bidding in Medicare Advantage: Effect of benchmark changes on plan bids." *Journal of health economics*, 32(6): 1301–1312.

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

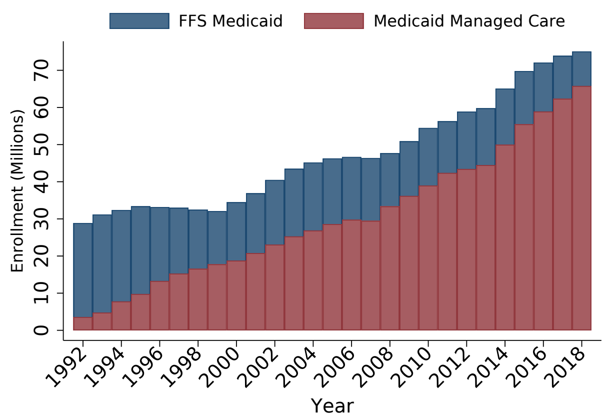
Spitz, Bruce. 2007. "Medicaid Agencies as Managed Care Organizations: An âĂIJActuarially SoundâĂİ Solution?" *Journal of health politics, policy and law*, 32(3): 379–413.

Wallace, Jacob. 2020. "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.

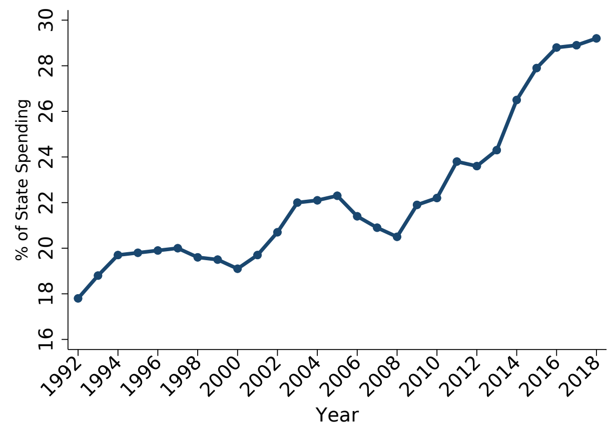
Zeckhauser, Richard. 1970. "Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives." *Journal of Economic theory*, 2(1): 10–26.

Figure 1: Trends in Medicaid Enrollment and Spending, 1992-2018

(a) Medicaid Enrollment in Fee-for-Service and Medicaid Managed Care

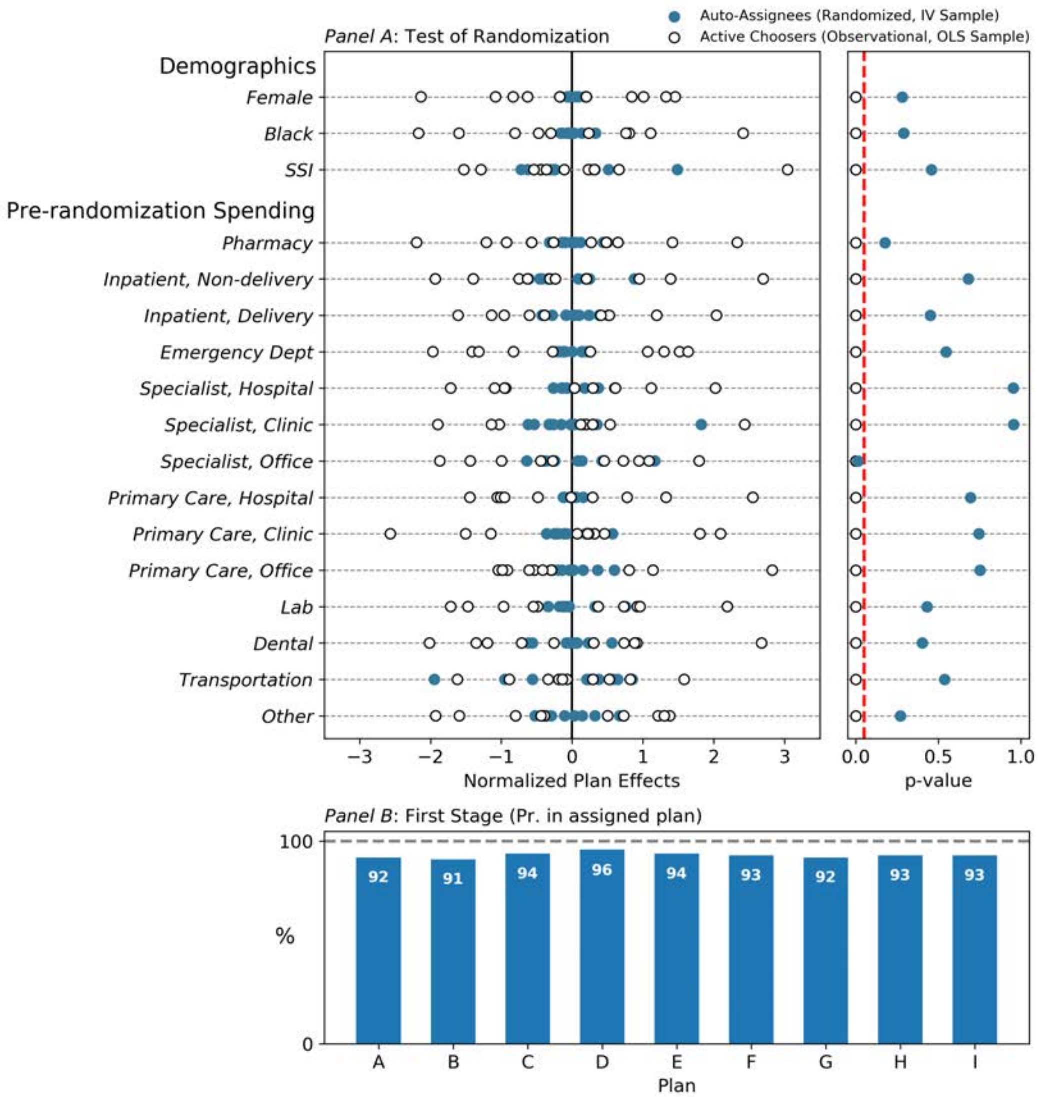


(b) Medicaid Spending as % of Total State Spending



Note: Figure displays trends in Medicaid Mangaed Care spending and enrollment for years 1992-2018. Spending data is pulled from the Medicaid and CHIP Payment and Access Commission's December 2020 report (Medicaid, Payment and Commission, 2020). State budget includes state and federal funds. Managed care enrollment counts come from several sources. Counts for years 1992-2000 are pulled from the Kaiser Commission on Medicaid and the Uninsured's December 2001 fact sheet (#2068-03) (Kaiser Family Foundation, 2001); 2001-2008 from the same commission's February 2010 policy brief (#8046) (Kaiser Family Foundation, 2010); 2009-2011 from the CMS' July 2011 Medicaid Managed Care Enrollment Report (Centers for Medicaid and Medicare Services, 2011); 2013 is pulled from a CMS and Mathematica July 2013 report (Centers for Medicaid and Medicare Services and Mathematica Policy Research, 2013); 2015 and 2018 from CMS and Mathematica Winter 2016 and 2020 policy reports (respectively) (Centers for Medicare and Medicaid Services, 2016; Centers for Medicaid and Medicare Services and Mathematica Policy Research, 2018). Enrollment counts for 2012, 2014, 2016, and 2017 are obtained using interpolation. Total enrollment counts are taken from the Medicaid and CHIP Payment and Access Commission's December 2020 report and FFS Medicaid enrollment is calculated as the difference between total Medicaid enrollment and managed care enrollment.

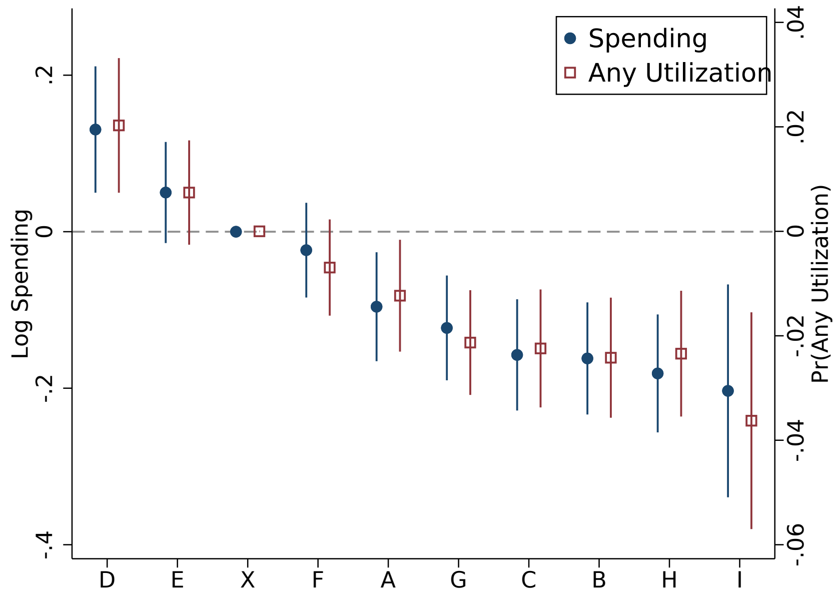
Figure 2: First Stage and Instrument Balance on Predetermined Characteristics



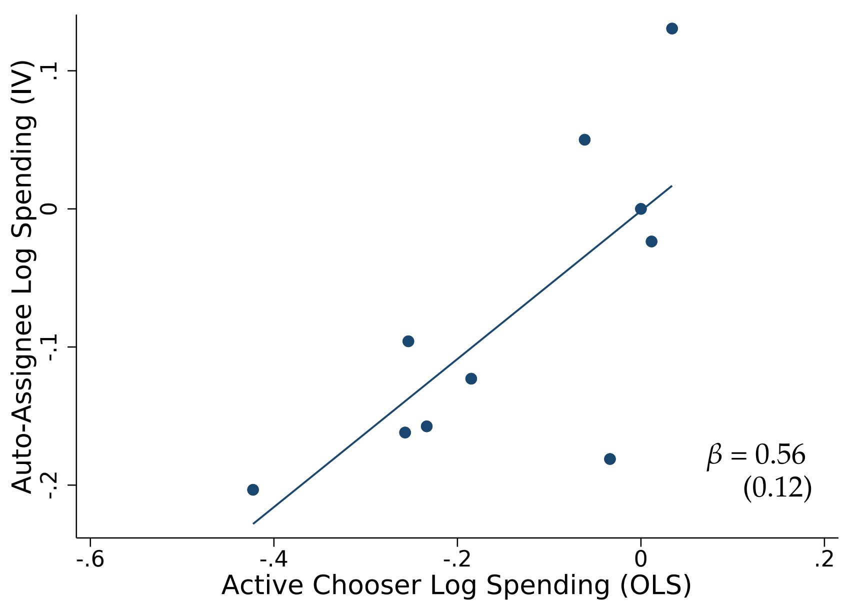
Note: Figure displays a balance test for the randomization in Panel (a) and first stage regression coefficients in Panel (b). Pre-determined characteristics include demographics and healthcare utilization in FFS Medicaid prior to randomized auto-assignment to a managed care plan. Each enrollee spent 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 A2. 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. 2), 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 A1 reports all first stage coefficients.

Figure 3: Main Results: IV Plan Effects on Healthcare Utilization

(a) Causal (IV) Effects on Spending and Utilization

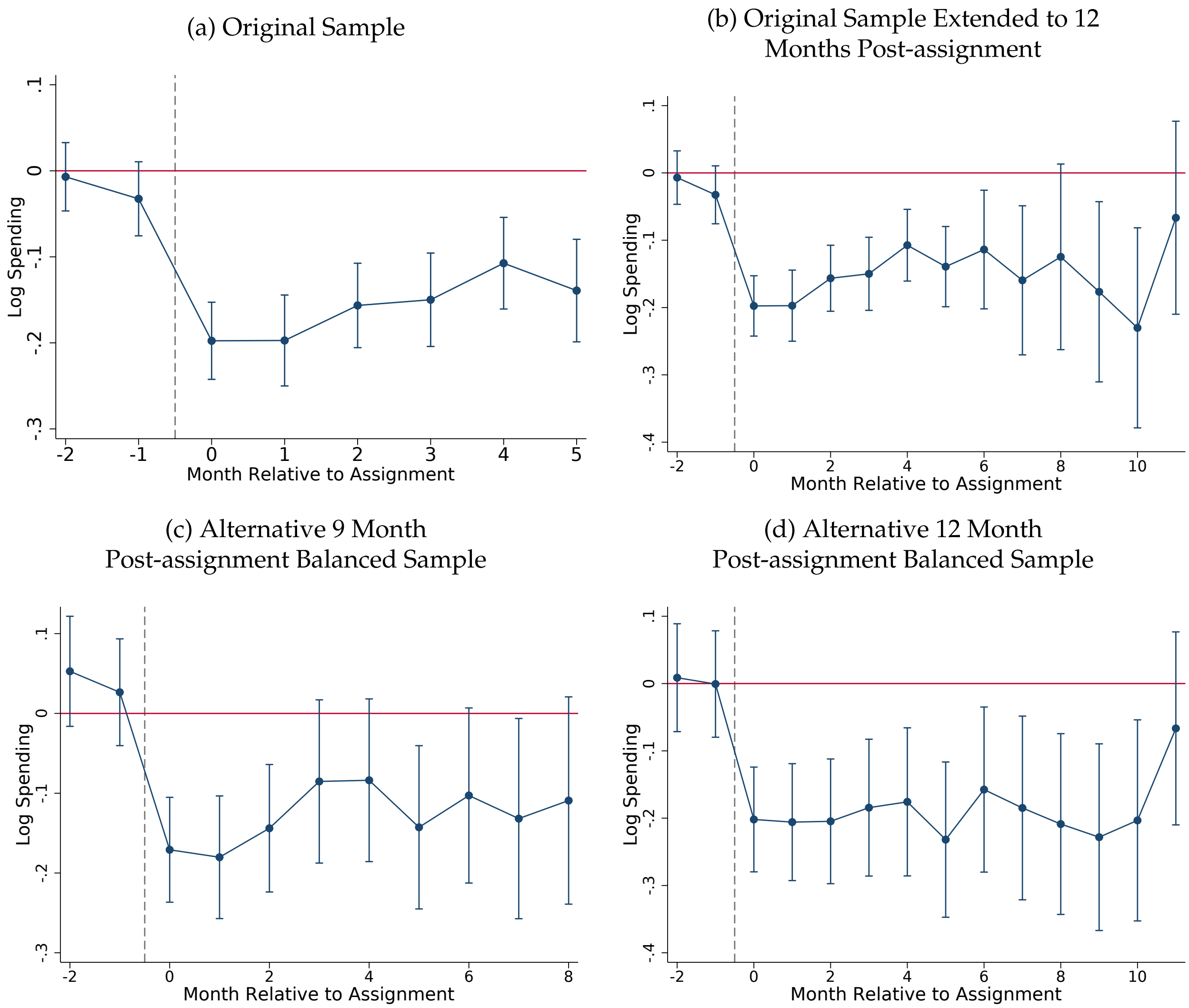


(b) Observational Versus Causal (IV) Spending Estimates



Note: Figure displays a main result of the paper—plan effects on healthcare utilization identified by random plan assignment. Panel (a) plots IV coefficients corresponding to Eq. 3, 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. 1. 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 2 and A7 for tabular forms of these results and for complete details on the control variables and reweighting.

Figure 4: Persistence: Effects By Time Since Assignment to a Plan



(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 A6, 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 (3), each point is estimated from a separate regression (one for each τ) of the form:

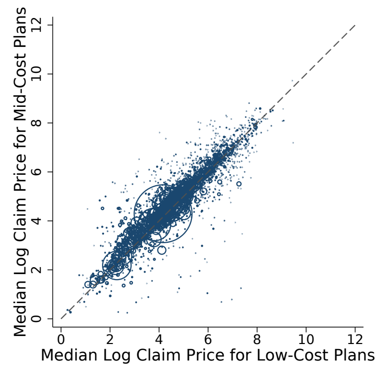
$$\log(\text{Spending} + 1)^{\tau}_{ict} = \alpha^{\tau} + \phi^{\tau}_{ct} + \delta^{\tau} X_{ict} + \lambda^{\tau}_{\text{low}} \mathbf{1}[\widehat{\text{Low Plan}}_{ict}] + \lambda^{\tau}_{\text{high}} \mathbf{1}[\widehat{\text{High Plan}}_{ict}] + \epsilon^{\tau}_{ict}.$$

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. None of the coefficients presented, including coefficients for $\tau = -1$ and $\tau = -2$, 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.

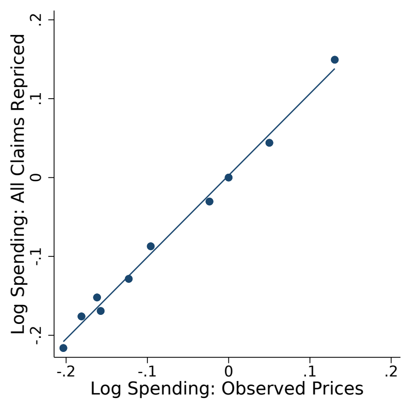
Figure 5: Transaction Price Differences Do Not Account for Spending Differences

(a) Inpatient Prices (DRGs)
Medium- vs. Low-Spending Plans

(b) Outpatient Prices (HCPCS)
Medium- vs. Low-Spending Plans

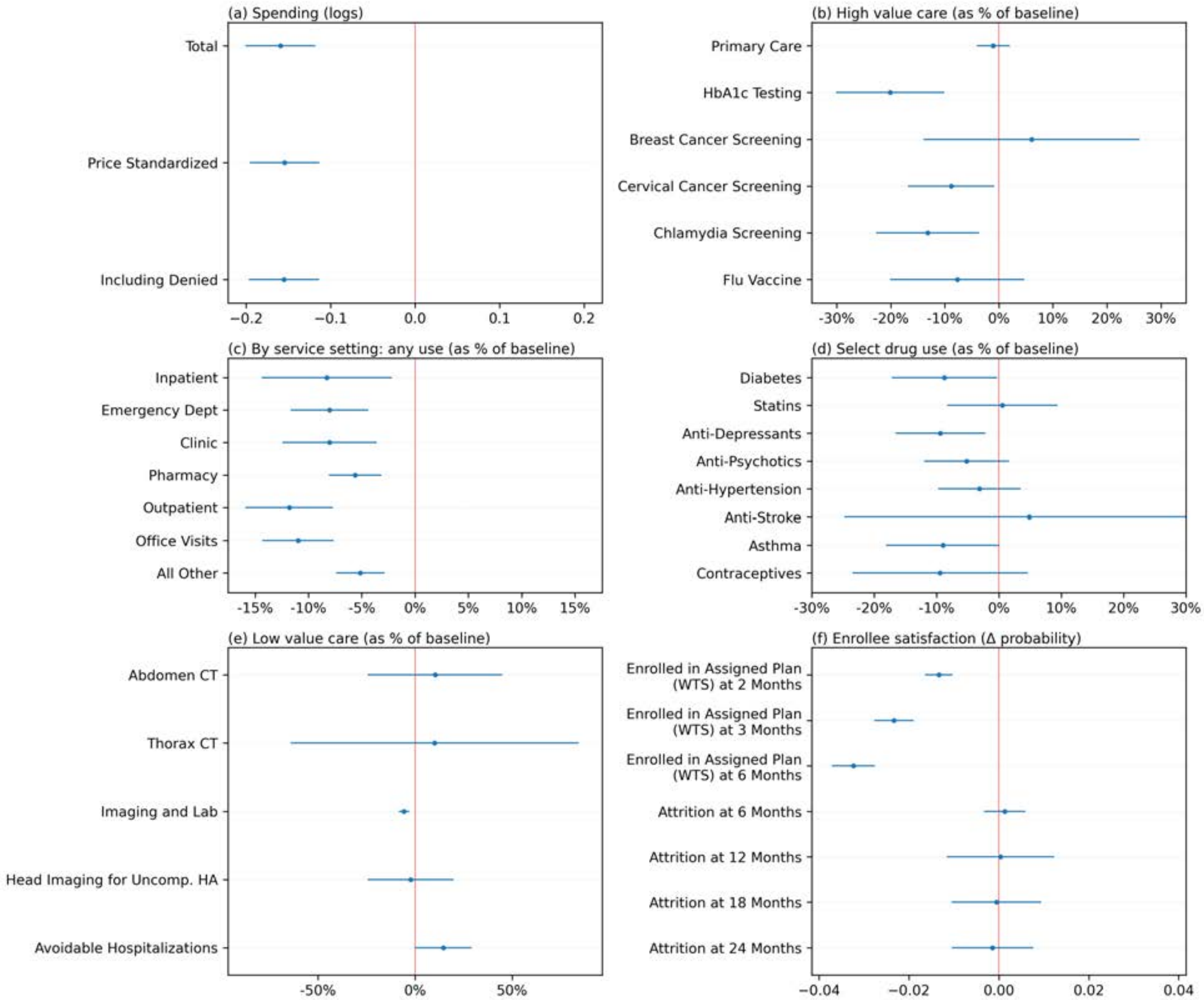


(c) Normalized Spending: Repricing All Claims to Common Price List



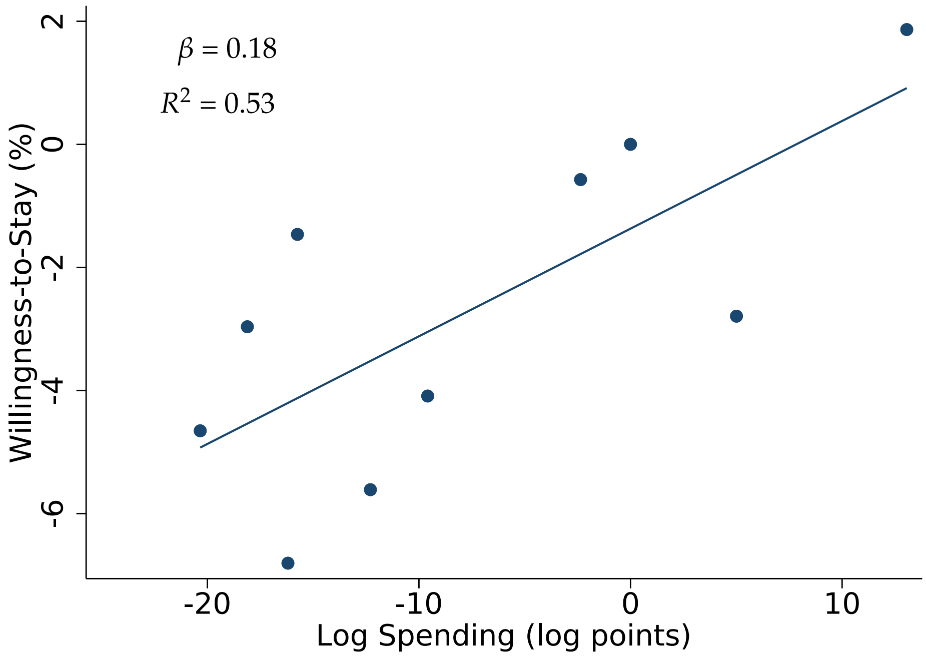
Note: Figure shows the minor role played by transaction prices in explaining spending differences across plans. The top two panels divide plans into high-, medium-, and low-spending groups as described in the text. We focus on medium- and low-spending plan groups as the high spender is a single plan 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 6: Low- Versus Medium-Spending Plan Effects Across Settings and Outcomes



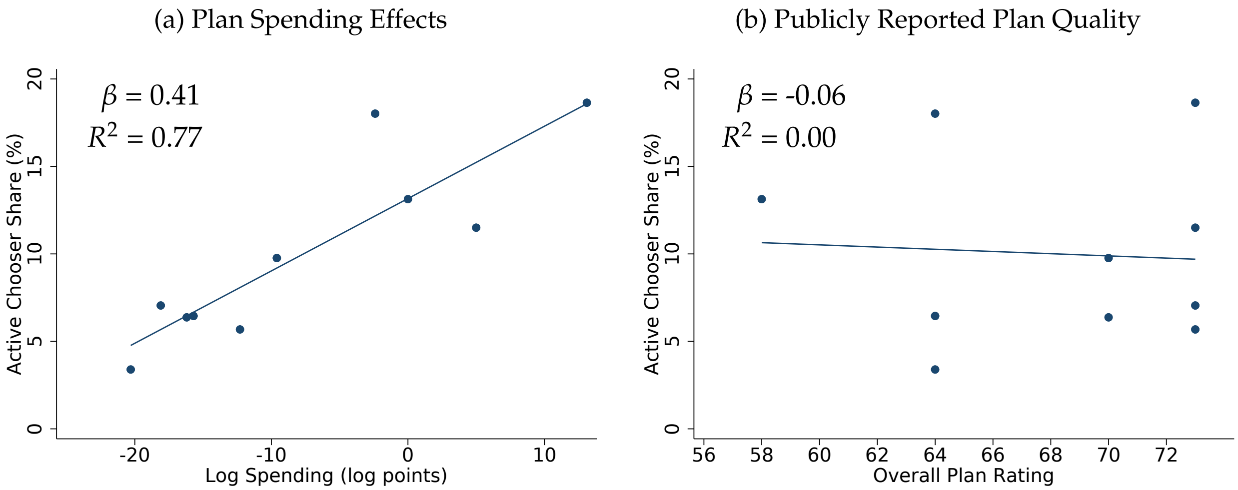
Note: Figure shows spending and utilization in low-spending plans compared to medium-spending plans across various categories and service settings. Plans are divided into three sets: low- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high- (Plan D) spending. We estimate a modified version of the IV regression in Eq. 3 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 A6 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 remain enrolled in the assigned plan (willingness-to-say; 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 A14, A15, A16, and A17.

Figure 7: Consumer Satisfaction Versus Plan Spending Effects



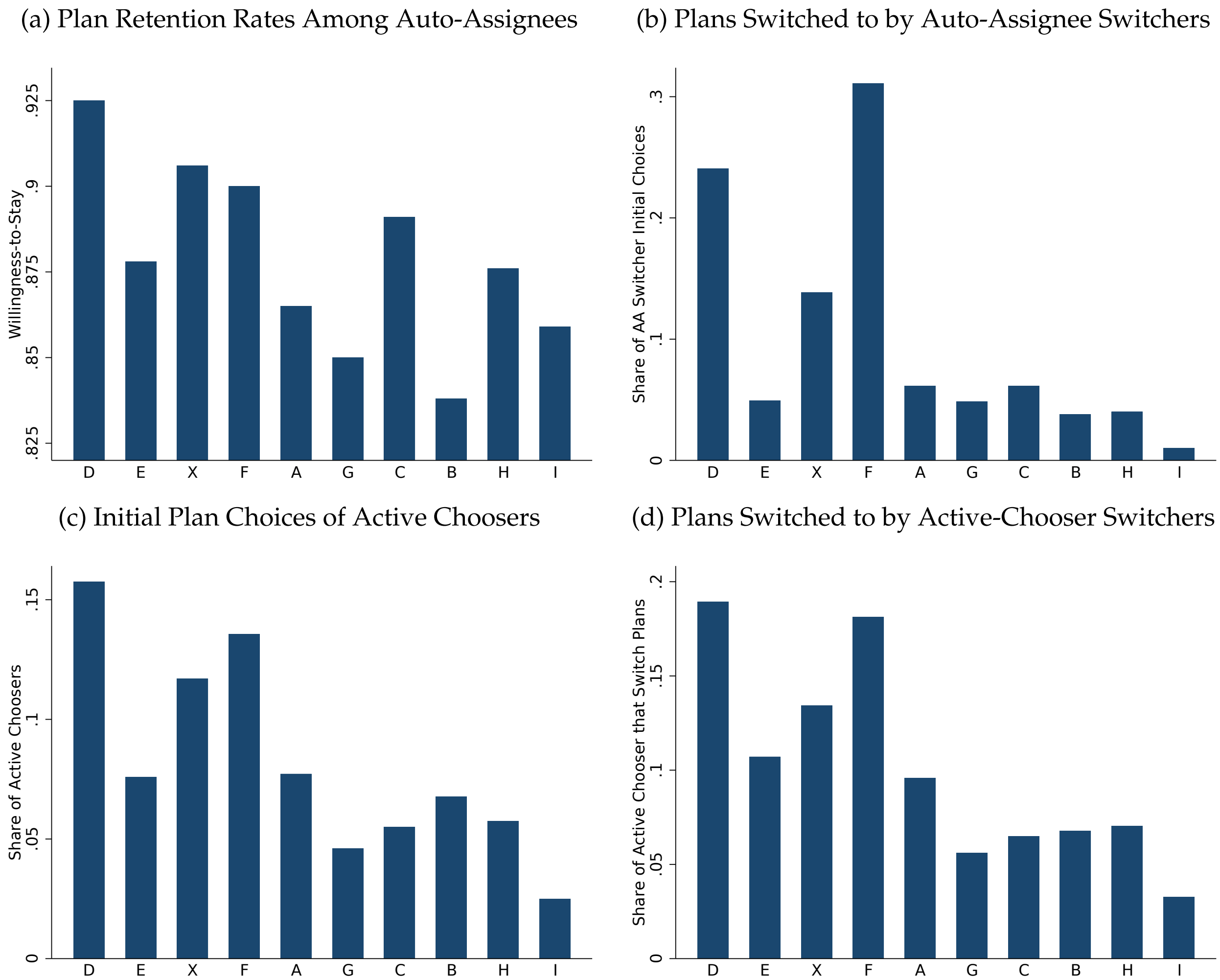
Note: Figure shows the strong correspondence between willingness-to-stay (WTS) and IV plan spending effects. WTS measures beneficiary satisfaction as the probability that a (randomly assigned) auto-assignee remains enrolled in their assigned plan through six months post-assignment. Each plan corresponds to one point, with the coordinates corresponding to the coefficient estimates from Table 2. The line-of-best-fit corresponds to the OLS fit of the 10 points, with the β and R^2 from that regression overlaid on the figure.

Figure 8: Plan Market Shares Versus Plan Spending Effects and Publicly Reported Plan Quality



Note: Figure shows the strong correspondence between active chooser share and plan spending (Panel a) and the weak correspondence between active chooser share and publicly reported plan quality (Panel b). Active chooser share is the percent of active choosers who initially chose the plan. Overall plan quality is measured by plan satisfaction as reported by each plan's enrollees. Each plan corresponds to one point in each panel. The plotted line corresponds to the OLS fit of the 10 points, with the β and R^2 from that regression overlaid on the figure.

Figure 9: The Highest Spending Plans Tend to Be Most Preferred



Note: Figure shows various measures of revealed preference for plans among different groups. Plans are ordered along each horizontal axis according decreasing causal effects on spending, with Plan D having the highest (most positive) impact on spending and Plan I have the lowest (most negative) impact. Panel (a) shows retention statistics among auto-assignees to their plan of random assignment, calculated as the 6-month willingness-to-stay coefficient estimate (Table 2) plus the mean retention rate. Panel (b) shows the plans chosen by auto-assignees who left their assigned plans. Panel (c) shows the initial plan choices of beneficiaries who made an active choice. And Panel (d) shows the second plan choices of these active choosers who switched plans after an initial choice.

Table 1: Summary Statistics

	Mean	Std. Dev	Observations
	(1)	(2)	(3)
Demographics			
Female (%)	40.1	49.0	393,570
White (%)	27.2	44.5	393,570
Black (%)	51.8	50.0	393,570
Age (years)	35.8	12.7	393,570
Healthcare Spending, $ per enrollee-month			
Total	509.74	2877.08	393,570
Office Visits	21.42	164.54	393,570
Clinic	52.50	280.33	393,570
Inpatient	220.04	2546.34	393,570
Outpatient	41.19	301.67	393,570
Emergency Dept.	15.80	99.58	393,570
Pharmacy	74.95	453.78	393,570
All Other	83.84	621.26	393,570
Drug Days Supply, days			
Diabetes	1.11	8.69	393,570
Statins	0.83	5.79	393,570
Anti-Depressants	1.31	7.80	393,570
Anti-Psychotics	1.49	8.64	393,570
Anti-Hypertension	1.32	7.91	393,570
Anti-Stroke	0.10	2.14	393,570
Asthma	0.46	4.11	393,570
Contraceptives	0.25	3.28	393,570
High-Value Care, per 1,000 enrollee-months			
HbA1c Testing	5.49	73.91	393,570
Breast Cancer Screening	1.47	38.29	393,570
Cervical Cancer Screening	7.29	85.05	393,570
Chlamydia Screening	6.61	81.01	393,570
Low-Value Care, per 1,000 enrollee-months			
Abdomen CT	0.33	18.17	393,570
Imaging and Lab	143.88	350.97	393,570
Head Imaging for Uncomp. HA	1.90	43.52	393,570
Thorax CT	0.09	9.43	393,570
Avoidable Hospitalizations	5.44	73.56	393,570

Note: Table reports summary statistics for the auto-assignee sample (used in the main IV analysis) over the first 6 months post-assignment. Observations are at the enrollee-month level. See Section 2.3 for details on the auto-assignee sample restriction and Appendix B for detailed descriptions of the low- and high-value care measures.

Table 2: Main Results: Plan Effects on Spending and Plan Switching

	Summary Statistics			Regression Results								
				OLS Spending			IV Spending			Willingness-to-Stay	
	Number of Auto-Assignees (IV Sample)	% of Active Choosers Selecting Plan	For-profit	Log Spending	Log Spending	Log Spending, Weighted	Log Spending	Log Spending	Any Spending in Enrollee-Month?	Enrolled in Assigned Plan at 3 mos?	Enrolled in Assigned Plan at 6 mos?
Plan	(1)	(2)		(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
D	2,886	18.64		0.200**	-0.065**	0.034	0.171**	0.131**	0.020**	0.014**	0.019**
				(0.026)	(0.023)	(0.052)	(0.050)	(0.041)	(0.007)	(0.004)	(0.005)
E	6,693	11.50		0.069*	-0.081**	-0.061	0.058	0.050	0.007	-0.025**	-0.028**
				(0.033)	(0.027)	(0.056)	(0.040)	(0.033)	(0.005)	(0.003)	(0.004)
X	8,758	13.13									
F	8,407	18.02		0.281**	-0.106**	0.012	-0.011	-0.024	-0.007	-0.001	-0.006
				(0.032)	(0.026)	(0.051)	(0.036)	(0.031)	(0.005)	(0.003)	(0.004)
A	8,401	9.76		-0.293**	-0.296**	-0.253**	-0.101*	-0.096**	-0.012*	-0.030**	-0.041**
				(0.038)	(0.031)	(0.059)	(0.042)	(0.035)	(0.005)	(0.004)	(0.004)
G	8,303	5.68		0.016	-0.207**	-0.185**	-0.134**	-0.123**	-0.021**	-0.041**	-0.056**
				(0.045)	(0.034)	(0.061)	(0.041)	(0.034)	(0.005)	(0.004)	(0.004)
C	6,202	6.45		0.027	-0.173**	-0.233**	-0.166**	-0.157**	-0.022**	-0.010*	-0.015**
				(0.036)	(0.031)	(0.065)	(0.044)	(0.036)	(0.006)	(0.004)	(0.005)
B	7,582	6.37	✓	-0.528**	-0.425**	-0.257**	-0.178**	-0.162**	-0.024**	-0.047**	-0.068**
				(0.043)	(0.030)	(0.070)	(0.042)	(0.036)	(0.006)	(0.004)	(0.004)
H	6,996	7.05	✓	-0.057	-0.086*	-0.034	-0.158**	-0.181**	-0.023**	-0.020**	-0.030**
				(0.041)	(0.034)	(0.064)	(0.046)	(0.038)	(0.006)	(0.004)	(0.005)
I	1,363	3.39	✓	-0.542**	-0.397**	-0.423**	-0.165+	-0.203**	-0.036**	-0.030**	-0.047**
				(0.046)	(0.041)	(0.089)	(0.084)	(0.069)	(0.011)	(0.006)	(0.008)
Mean (spend displayed in dollars)				466.202	466.202	462.263	509.740	509.740	0.349	0.930	0.906
County x Year x Month FEs				X	X	X	X	X	X	X	X
Person-Level Controls					X	X		X	X	X	X
First Stage F-Statistic							7143.134	7030.613	7030.613		
Obs: Enrollees										65592	65592
Obs: Enrollee X Months				592153	592153	392094	393570	393570	393570		

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 3–8, plan regressors correspond to the plan of current enrollment in the enrollee-month. For the IV regressions (columns 6–8), these are instrumented with plan of initial assignment. Kleibergen-Paap F statistics from the first stage are reported. See Table A1 for first stage coefficients. In columns 9 and 10, 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 8 and enrollees in columns 9 and 10. OLS regressions include only active-choosers; see Table A18 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 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

[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]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)

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

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 and 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 D.2.

C Robustness and Alternative Specifications

C.1 Plan Group IV Regressions

Some of our results involve 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:

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

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:

$$\text{Low Plan}_{ict} = \phi_{1ct} + \delta_1 X_{ict} + \sum_{j \in \text{High,Low}} \lambda_{1j} \mathbf{1}[\text{Assigned j}_{\,ict}] + \eta_{1,ict},$$
$$\text{High Plan}_{ict} = \phi_{2ct} + \delta_2 X_{ict} + \sum_{j \in \text{High,Low}} \lambda_{2j} \mathbf{1}[\text{Assigned j}_{\,ict}] + \eta_{2,ict}.$$

Results using this specification are presented in Figure 6 and Tables A6 and A19.

C.2 Unfolding of Effects Over Event Time

In Table A5, 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.

In Table A8 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 2 and Figure 6 persist into the longer run—for the minority of managed care enrollees that have these longer enrollment spells.

D Mechanisms

D.1 Selection and organization of providers

We briefly explore provider 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.

D.1.1 Measuring provider network breadth

We measure network breadth as the share of simulated physician and hospital visits from a given zip code covered by each plan's network in each year. To simulate physician and hospital visits, we use estimates from models of physician and hospital demand in Wallace (2020), 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 for each plan (i.e., an unconstrained counterfactual). 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. Because there is substantial variation between plans in the out-of-network (OON) hassle costs in the hospital demand model, we also assess the correlation between our causal plan effects and plan-specific OON hassle costs.[5]

D.1.2 Assessing whether provider network breadth mediates our causal plan effects

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 Tables A10 and A11 presents the results of this analysis for log spending. Column 1 of Appendix Table A10 reproduces our primary estimate of the causal effects of assignment to a low- or high-spending plan on log spending. In Column 2 we additionally instrument for provider network breadth using the network breadth (which varies at the plan-by-zip level) of the plan of assignment and the enrollee's zip code. The additional control for network breadth does not change the large estimated differences across plan spending groups. In Appendix Table A11 we present an analogous exercise for the estimation of the individual plan effects (rather than plan spending groups). Additionally instrumenting for provider network breadth does not substantially reduce the dispersion of the plan effects as measured by their range or standard deviation. In Appendix Tables A12 and A13 we demonstrate that our consumer satisfaction results are also not sensitive to including instruments for provider network breadth in the model.

Second, we plot our causal estimates of plan-level spending and willingness-to-stay (i.e., consumer satisfaction) against plan-level measures of provider network breadth. We plot these relationships in Appendix Figure A8. Panel (a) plots our estimated plan-level spending effects against provider network breadth measured at the plan-level. The slope of the line of best fit if we include all plans is close to zero. This can be reconciled with evidence from Gruber and McKnight (2016) and Wallace (2020) that narrower provider networks reduce 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. This "provider-owned" 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 (2020) to separately identify the effect of networks from other dimensions of MMC plans.

If we exclude the provider-owned plan and plot the line of best fit for log spending through the 9 plans, we observe a steeper slope and tighter fit (i.e., narrower network plans have lower spending) but the relationship is noisy ($R^2 = 013$). We observe a stronger linear relationship between our estimated plan-level willingness-to-stay effects against provider network breadth in the rightmost figure of Panel (a), though that relationship is also noisy given the small number of points. However, if we exclude the provider-owned plan, we observe a statistically significant relationship between willingness-to-stay and provider network breadth, with the variation in plan-level network breadth explaining 48% of the between-plan differences in willingness-to-stay.

Because plans differ in how difficult they make it for enrollees to seek out-of-network (OON) care, we also assessed the relationship between our causal plan effects and estimates of the OON "hassle costs" for each plan in Panel (b). These estimates are based on a model of hospital demand in Wallace (2020), which includes a "hassle cost" for going to an out-of-network provider, are not

[5] Additional details on network breadth measure construction and summary statistics are available in Wallace (2020).

rescaled (i.e., they are the values recovered by the conditional logit model of hospital choice). We find no relationship between the OON hassle costs and plan spending (with or without the provider-owned plan). However, if we exclude the provider-owned plan we do see some suggestive evidence that enrollees randomly assigned to plans with larger hassle costs (i.e., more negative values on the x-axis) have lower satisfaction. We find that OON hassle costs explained 24% of the variation in WTS in the rightmost figure of Panel (b).

D.1.3 Assessing whether provider steering mediates our causal plan effects

Another aspect of provider networks, beyond the question of broad versus narrow, involves whether certain plans steer patients to providers with more efficient practice styles (Glied, 2000). To assess this, we used administrative data on enrollees' healthcare utilization in the post-assignment period to attribute them to the physician or hospital with whom they utilized the most services. A limitation of this approach is that enrollees with no physician or hospital spending could not be attributed to a provider. This is particularly problematic in our setting given evidence that random plan assignment impacts the extensive margin of whether enrollees use any physician or hospital care.

With the caveat that we are conditioning on a pair of post-treatment outcomes (enrollees non-random choice of provider for a sample limited to enrollees with physician and hospital claims), column 3 of Appendix Table A10 reproduces our primary estimates for the sub-sample of enrollees we were able to attribute to a physician or hospital. Relative to the estimated effect of assignment to a low-cost plan in the full sample, the effect size in this sub-sample was smaller (-0.125), but remained highly statistically significant. Columns 4 and 6 demonstrate that the estimated spending difference was attenuated by 15-18% when we controlled for provider fixed effects. We observed similar reductions in the magnitude of the dispersion of individual plan effects when we controlled for provider fixed effects (Appendix Table A11). Hence, while the analysis is complicated by sample selection and the nonrandom sorting of enrollees to providers, the results suggest that steering to more efficient providers partly mediates our causal plan effects on spending, but does not appear to explain plan effects on consumer satisfaction (Appendix Tables A12 and A13).

D.2 Re-pricing of Claims

In Section 5.1, 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 prices. The estimating equation is:

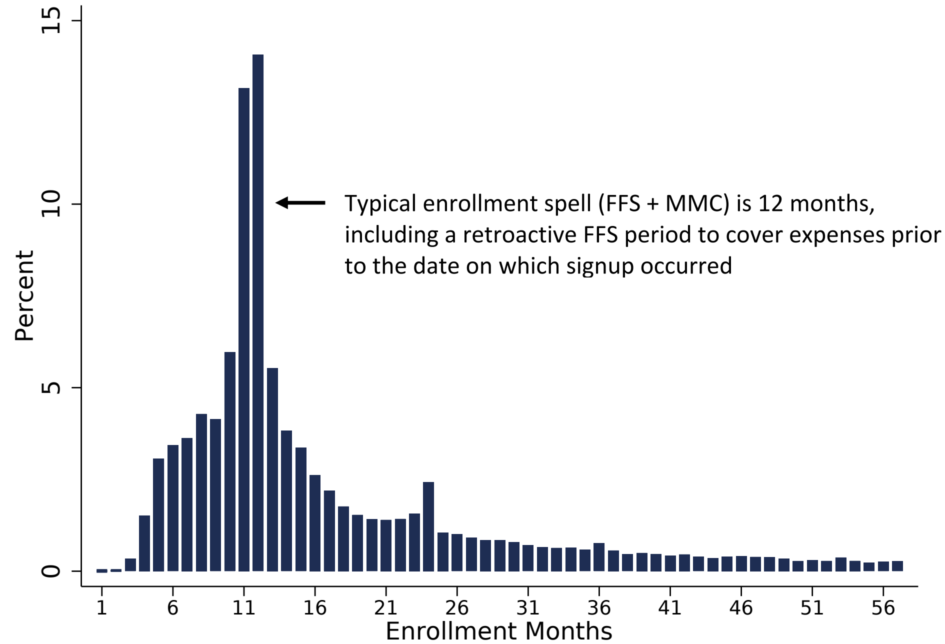
$$P_{djc} = \nu_d + \sum_{j=1}^{9} \Psi_j \mathbf{1}[\text{Plan j }] + \mu_{djc},$$

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 code-by-service group (physician office visit, ED visit, etc.) 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.

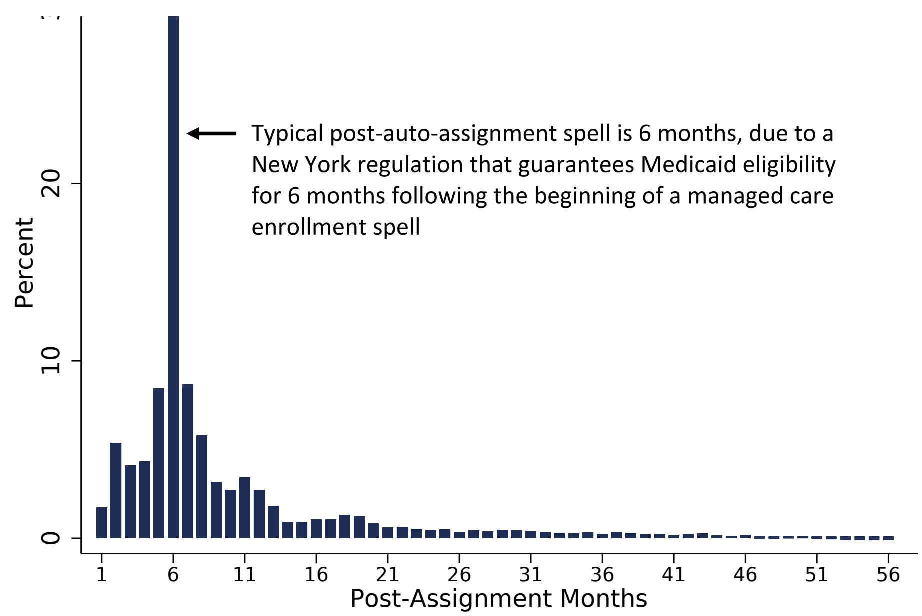
E Additional Figures and Tables

Appendix Figure A1: Enrollment Spell Lengths of Auto-Assignees

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

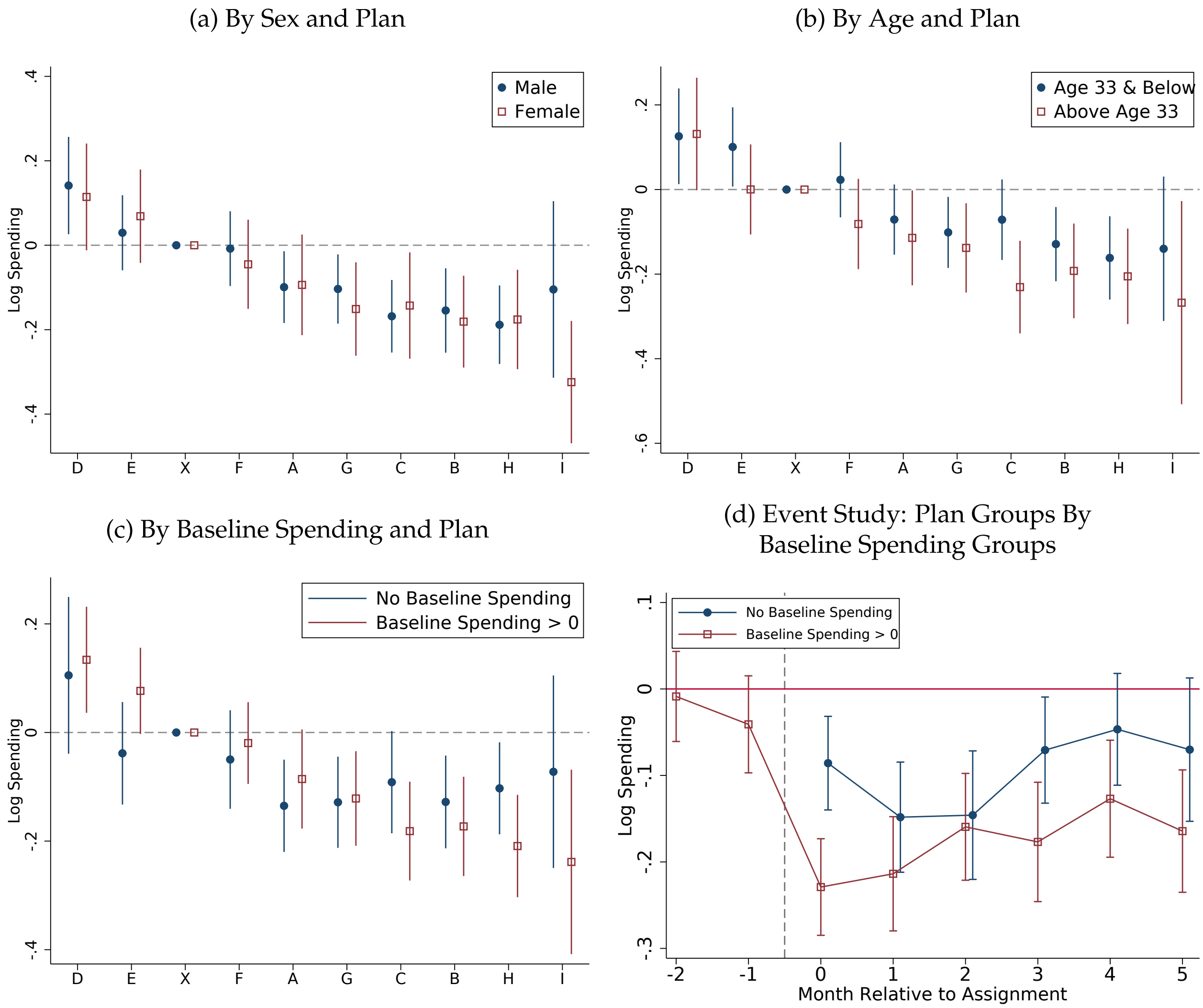


(b) Post-Auto-Assignment (MMC) Spell Lengths



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.

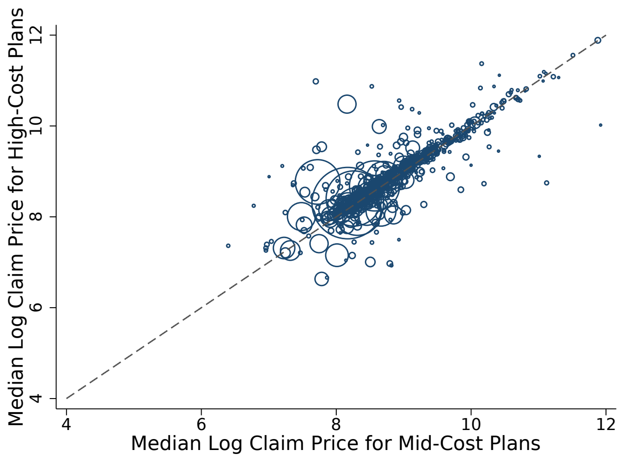
Appendix Figure A2: Plan Effects by Groups



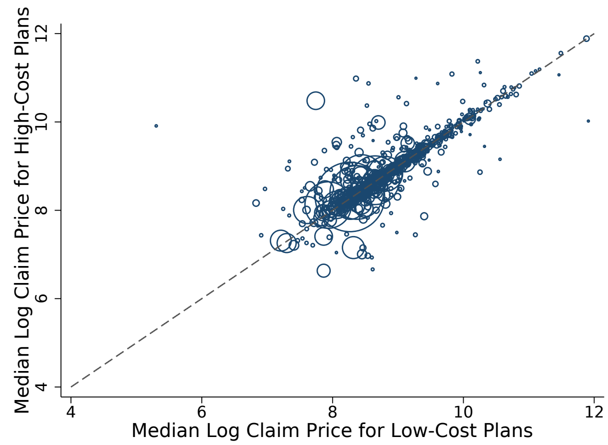
Note: Panels (a) through (c) replicate panel (a) of Figure 3 in various subsamples. These panels plot IV coefficients corresponding to Eq. 3, where the dependent variable is log(total healthcare spending +1), estimated separately within the indicated subsamples. Plans are ordered identically to Figure 3. Standard errors are clustered at the county $\times$ year $\times$ month-of-assignment level. This is the level at which the randomization operates. Panel (d) replicates panel (a) of Figure 4 separately in subsamples defined by baseline spending.

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

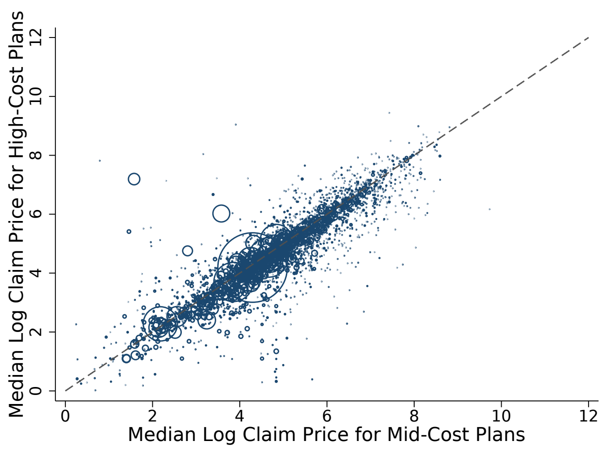
(a) Prices (DRGs) for High- vs. Medium-Spending Plans



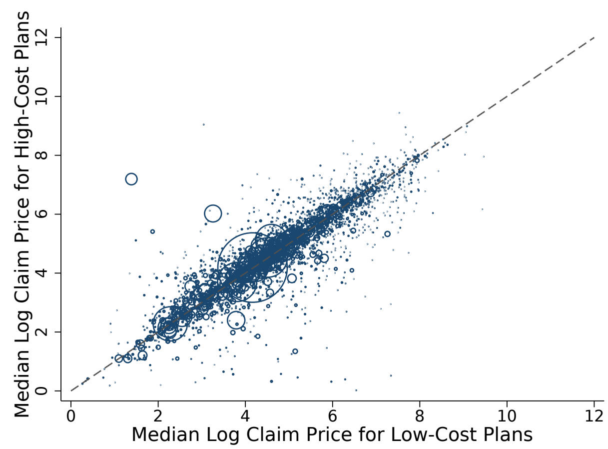
(b) Prices (DRGs) for High- vs. Low-Spending Plans



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

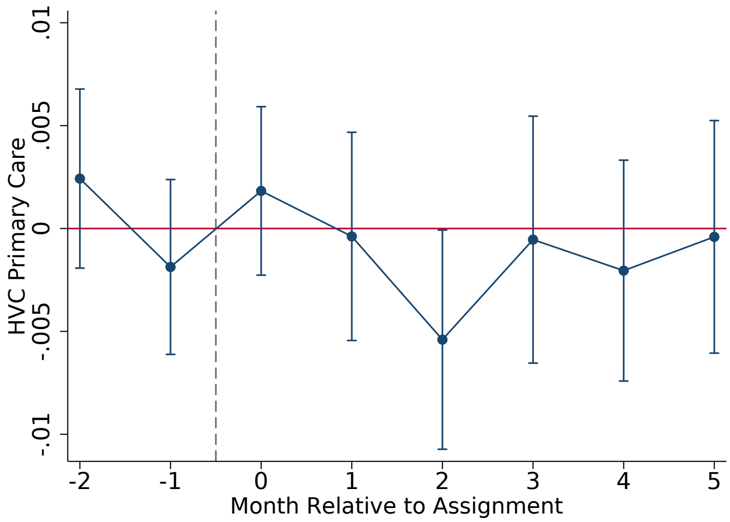


(d) Prices (HCPCS) for High- vs. Low-Spending Plans



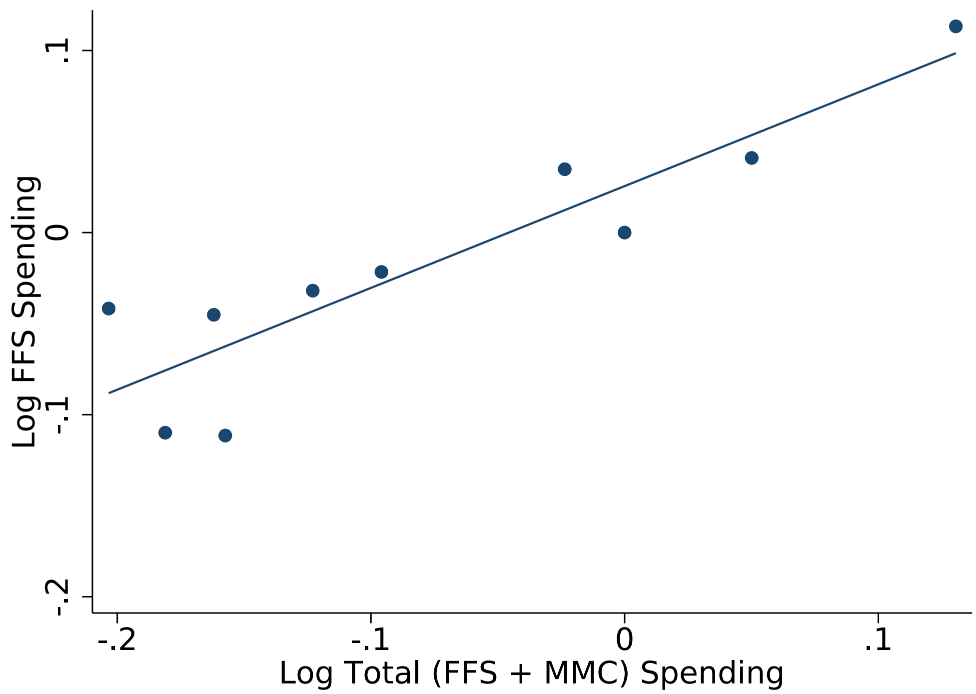
Note: Figure compares prices for inpatient admissions and outpatient services between high-, medium-, and low-spending plans. We divide plans into high-, medium-, and low-spending groups as described in the text. Figure 5 shows analogous comparisons for low- versus medium-spending plans. Each circle 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 5 for additional notes.

Appendix Figure A4: Primary Care Use by Time Since Plan Assignment



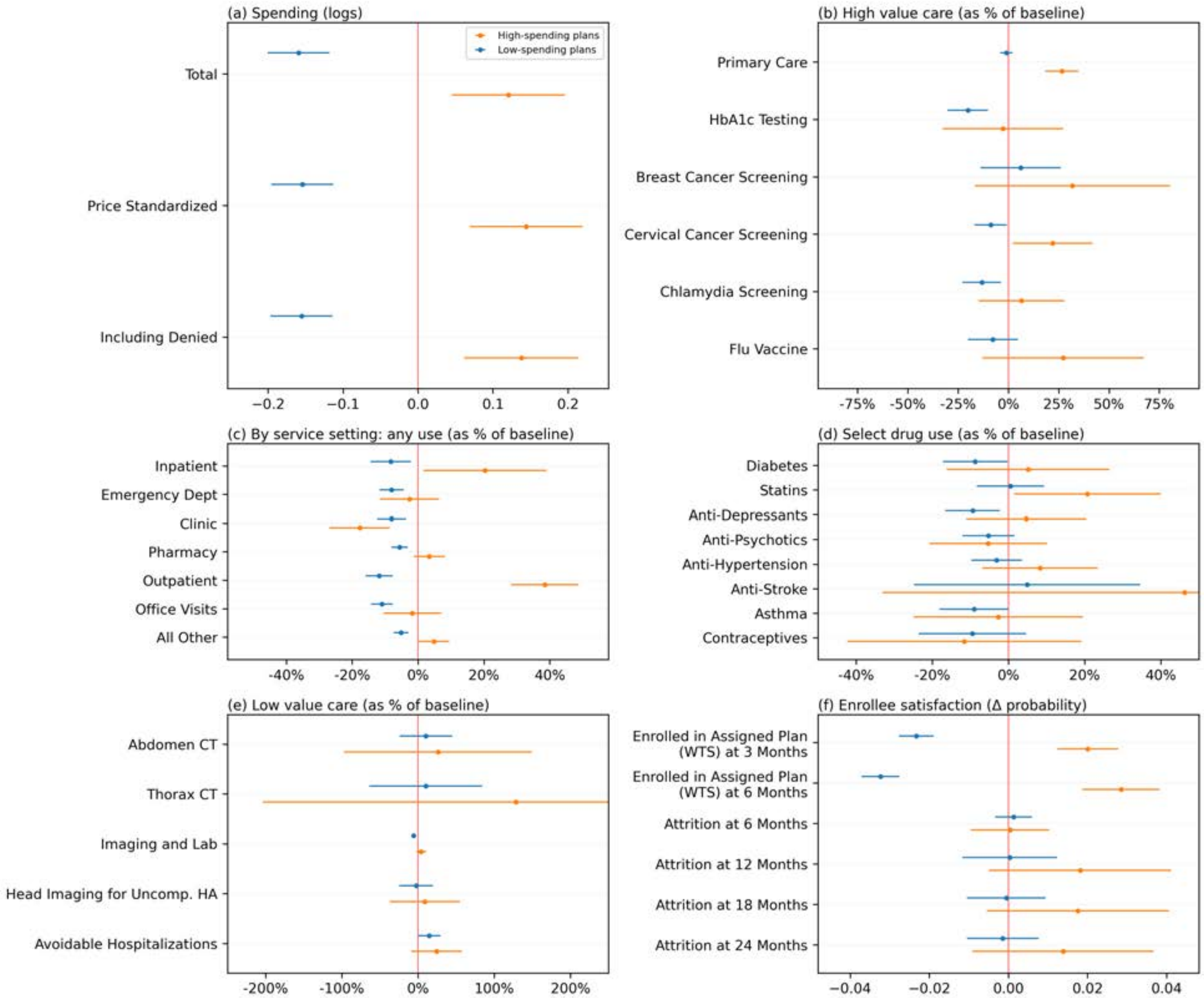
Note: Figure displays results in the spirit of difference-in-difference event studies showing the spending impacts of being assigned to a low- versus medium-spending plan. The specification follows Figure 4 but uses an indicator for any primary care physician (PCP) visit in the enrollee-month as the dependent variable. See Figure 4 for additional detail.

Appendix Figure A5: Carved-Out FFS Claims versus Total (FFS and MMC) Claims



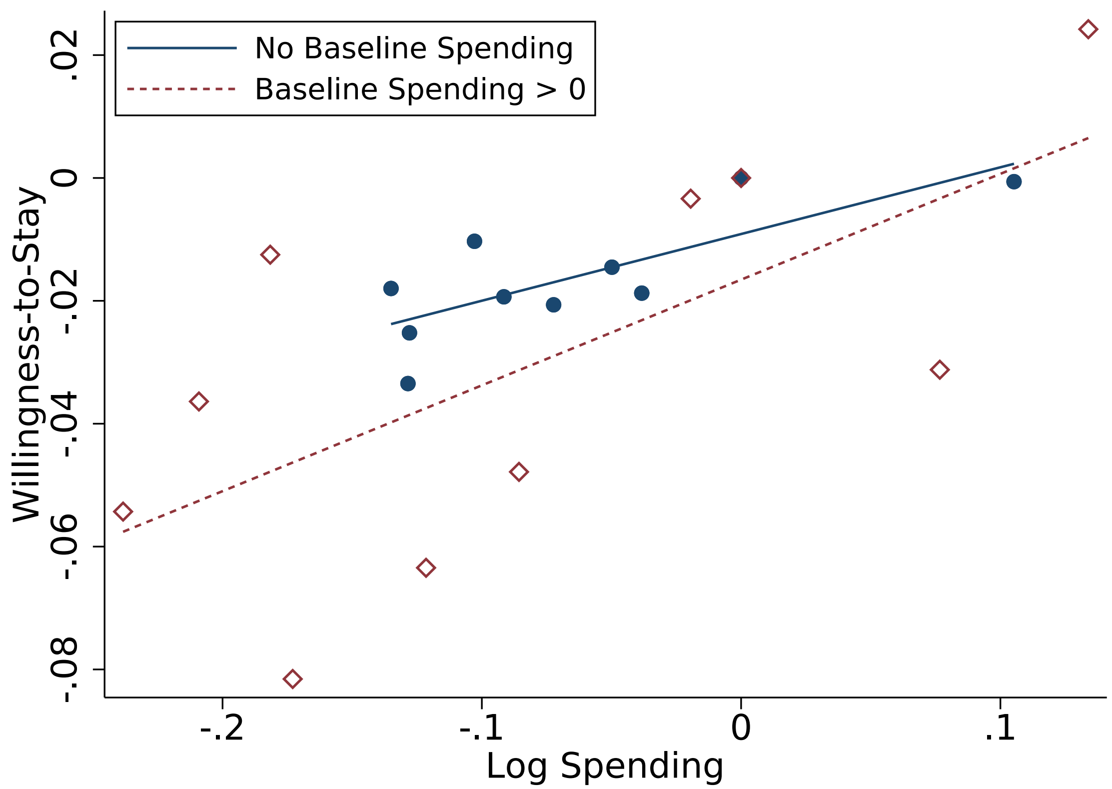
Note: Figure plots IV estimates of plan effects on carved-out FFS claims against plan effects on total (FFS plus MMC) claims. The sample is the main IV analysis sample. Carved-out FFS claims for MMC enrollees are paid and reported directly by the state, rather than by the plans despite occurring during MMC enrollment. Markers correspond to plans. The coefficients plotted along the horizontal axis are identical to those reported in Figure 5. 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.

Appendix Figure A6: Extending the Figure 6 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 6 notes for additional detail.

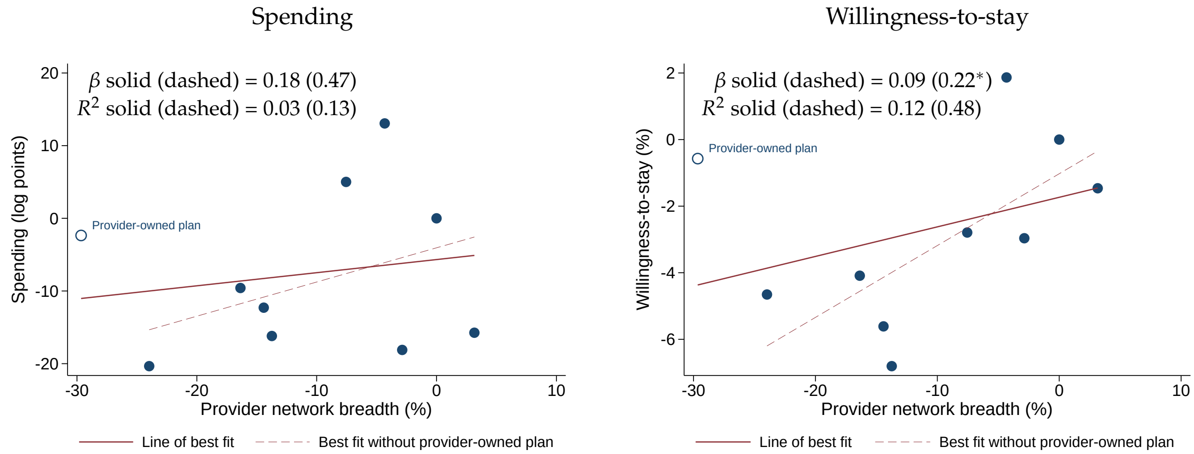
Appendix Figure A7: Auto-Assignees Divided by Baseline (Pre-Assignment) Spending



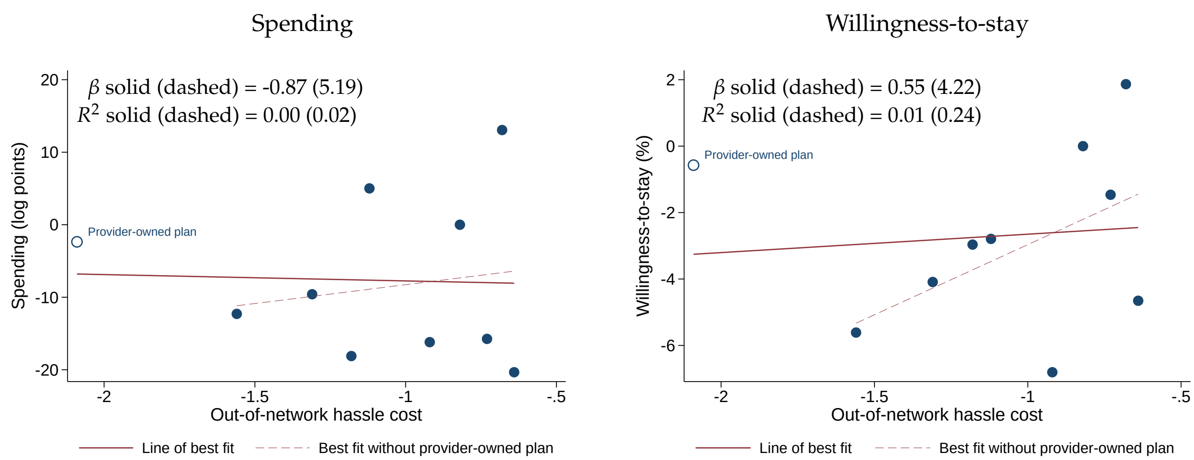
Note: Figure shows the correspondence between willingness-to-stay (WTS) and IV plan spending effects. In the top panel, each plan corresponds to one point, with the coordinates corresponding to the coefficient estimates from Table 2. In the bottom panel, each plan corresponds to two points: The WTS and 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 A8: Association Between Plan Effects and Provider Network Characteristics

(a) Provider Network Breadth



(b) Out-Of-Network Hassle Costs



Notes: Figure displays the association between the main results of the paper—causal plan effects on healthcare spending and satisfaction—and plan-level measures of provider network breadth and out-of-network hassle costs. The leftmost figure in Panel (a) plots IV coefficients corresponding to Eq. (3), 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. The rightmost figure in Panel (a) presents analogous estimates for our willingness-to-stay measure on the y-axis, where the dependent variable is an indicator for whether an enrollee remained in their assigned plan at six months post-assignment, as in the last column of 2. The x-axis is identical in the two figures. In Panel (b), the plan effects on the y-axes are identical to Panel (a), but the x-axis now contains the plan-specific, out-of-network hassle cost from a model of hospital demand in (Wallace, 2020). Appendix Section D.1 describes how network breadth and out-of-network hassle costs are measured.

Appendix Table A1: 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	0.905**	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	0.933**	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	393570	393570	393570	393570	393570	393570	393570	393570	393570

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Note: Table reports coefficients from the nine first stage regressions defined in Equation 2. 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 $\times$ year $\times$ month-of-assignment and the count of months since plan assignment, both as indicators. Person level controls, as described in Table 2 are included as well in all columns. 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. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Appendix Table A2: 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	84.7	0
Black	1.2	.28	130.7	0
SSI	1.1	0.35	21.1	0
Other	1.2	0.32	33.6	0
Dental	1.1	0.40	44.6	0
Transportation	0.9	0.56	9.2	0
Lab	1.0	0.42	71.0	0
Pharmacy	1.4	0.18	43.1	0
Inpatient, Non-delivery	0.7	0.72	42.5	0
Inpatient, Delivery	1.0	0.41	41.4	0
Emergency Dept	0.9	0.52	112.6	0
Specialist, Hospital	0.4	0.95	37.9	0
Specialist, Clinic	0.4	0.95	23.0	0
Specialist, Office	2.4	0.01	53.1	0
Primary Care, Hospital	0.7	0.71	325.2	0
Primary Care, Clinic	0.6	0.79	65.0	0
Primary Care, Office	0.7	0.74	18.2	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 2. 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 A3: Summary Statistics

	Active Choosers		Auto-Assignees	
	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.587	0.492	0.401	0.490
White	0.340	0.474	0.272	0.445
Black	0.303	0.460	0.518	0.500
Age	34.4	12.7	35.8	12.7
Healthcare Spending:				
Total	461.29	2060.29	509.74	2877.08
Office Visits	65.64	458.22	21.42	164.54
Clinic	23.45	152.27	52.50	280.33
Inpatient	180.15	1681.76	220.04	2546.34
Outpatient	42.47	272.75	41.19	301.67
Emergency Dept.	9.92	63.84	15.80	99.58
Pharmacy	56.68	279.93	74.95	453.78
All Other	82.99	533.29	83.84	621.26
Drug Days Supply:				
Diabetes	1.61	10.92	1.11	8.69
Statins	1.32	7.71	0.83	5.79
Anti-Depressants	0.82	6.14	1.31	7.80
Anti-Psychotics	0.57	5.22	1.49	8.64
Anti-Hypertension	1.54	8.72	1.32	7.91
Anti-Stroke	0.07	1.83	0.10	2.14
Asthma	0.42	3.91	0.46	4.11
Contraceptives	0.60	5.20	0.25	3.28
High-Value Care:				
HbA1c Testing	0.0095	0.0968	0.0055	0.0739
Breast Cancer Screening	0.0050	0.0708	0.0015	0.0383
Cervical Cancer Screening	0.0246	0.1550	0.0073	0.0851
Chlamydia Screening	0.0142	0.1182	0.0066	0.0810
Low-Value Care:				
Abdomen CT	0.0007	0.0257	0.0003	0.0182
Imaging and Lab	0.2388	0.4264	0.1439	0.3510
Head Imaging for Uncomp. HA	0.0023	0.0481	0.0019	0.0435
Thorax CT	0.0001	0.0105	0.0001	0.0094
Avoidable Hospitalizations	0.0014	0.0374	0.0054	0.0736
Observations	592153		393570	

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 A16 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 A4: 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	-0.096**	-0.105**	-22.866	-0.012*	-0.087*	-19.468
	(0.035)	(0.039)	(13.965)	(0.005)	(0.035)	(12.342)
B	-0.162**	-0.179**	-24.168	-0.024**	-0.152**	-21.499
	(0.036)	(0.040)	(14.960)	(0.006)	(0.036)	(13.254)
C	-0.157**	-0.173**	-39.792**	-0.022**	-0.169**	-45.697**
	(0.036)	(0.040)	(14.696)	(0.006)	(0.036)	(12.711)
D	0.131**	0.145**	64.771**	0.020**	0.149**	62.200**
	(0.041)	(0.045)	(22.495)	(0.007)	(0.041)	(20.261)
E	0.050	0.055	13.791	0.007	0.044	7.680
	(0.033)	(0.036)	(14.641)	(0.005)	(0.033)	(13.005)
F	-0.024	-0.028	-2.421	-0.007	-0.030	-12.690
	(0.031)	(0.034)	(15.360)	(0.005)	(0.030)	(13.313)
G	-0.123**	-0.137**	6.205	-0.021**	-0.128**	0.355
	(0.034)	(0.037)	(15.377)	(0.005)	(0.034)	(13.793)
H	-0.181**	-0.197**	-52.253**	-0.023**	-0.176**	-49.695**
	(0.038)	(0.042)	(15.139)	(0.006)	(0.038)	(13.202)
I	-0.203**	-0.228**	-14.220	-0.036**	-0.216**	-25.466
	(0.069)	(0.076)	(28.463)	(0.011)	(0.069)	(25.168)
Mean	2.09	2.33	416.45	0.35	2.09	385.56
Observations	393570	393570	393570	393570	393570	393570

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

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 3. 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 5 Panel c and is described in full detail in Appendix D.2. 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 $\times$ year $\times$ month-of-assignment and the count of months since plan assignment, both as a set of indicators. Person level controls, as described in Table 2 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. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A5: 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	-0.255**	-0.274**	-183.437^{+}	-0.073*	-0.230**	-181.636*
	(0.054)	(0.059)	(100.052)	(0.033)	(0.054)	(91.806)
B	-0.345**	-0.371**	-211.016^{+}	-0.151**	-0.305**	-196.694^{+}
	(0.056)	(0.061)	(107.262)	(0.036)	(0.056)	(101.805)
C	-0.230**	-0.246**	-307.084**	-0.134**	-0.247**	-359.900**
	(0.054)	(0.058)	(92.780)	(0.034)	(0.053)	(86.369)
D	0.183*	0.198*	510.058**	0.122**	0.223**	501.853**
	(0.076)	(0.083)	(163.673)	(0.039)	(0.076)	(156.484)
E	0.099^{+}	0.107^{+}	102.790	0.053^{+}	0.092^{+}	44.021
	(0.055)	(0.059)	(106.447)	(0.031)	(0.054)	(97.910)
F	-0.009	-0.009	-46.962	-0.042	-0.011	-138.228
	(0.053)	(0.058)	(106.735)	(0.028)	(0.053)	(96.710)
G	-0.259**	-0.281**	20.656	-0.127**	-0.256**	-1.173
	(0.054)	(0.058)	(101.082)	(0.031)	(0.053)	(96.304)
H	-0.338**	-0.361**	-324.523**	-0.140**	-0.317**	-370.881**
	(0.059)	(0.064)	(104.479)	(0.037)	(0.058)	(96.730)
I	-0.333**	-0.365**	-47.533	-0.218**	-0.353**	-75.183
	(0.112)	(0.122)	(193.764)	(0.063)	(0.111)	(183.638)
Mean	4.41	4.84	2732.90	2.09	4.40	2602.14
Observations	65595	65595	65595	65595	65595	65595

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

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 3. The columns vary the parameterization of spending used as the dependent variable, as indicated in the column headers. The difference here compared to Table A4 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 A4 notes for additional details. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A6: 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	0.120**	0.135**	59.317**	0.020**	0.144**	62.798**
	(0.038)	(0.042)	(22.017)	(0.006)	(0.038)	(19.938)
Low-Cost Plans	-0.159**	-0.175**	-30.028**	-0.022**	-0.154**	-25.565**
	(0.021)	(0.023)	(9.081)	(0.003)	(0.021)	(7.988)
Mean	2.09	2.33	416.45	0.35	2.09	385.56
Observations	393570	393570	393570	393570	393570	393570

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Note: Table reports IV estimates of each plan grouping's causal effect on utilization, using a modification to the IV regression in Equation 3. 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. (4) in Section 3.1. Specifications otherwise follow Table A4. See Table A4 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	-0.308**	-0.265**	-0.296**	-0.253**	-0.043**	-0.019	-0.022
	(0.028)	(0.064)	(0.031)	(0.059)	(0.005)	(0.014)	(0.014)
B	-0.509**	-0.414**	-0.425**	-0.257**	-0.055**	-0.068**	-0.066**
	(0.031)	(0.084)	(0.030)	(0.070)	(0.005)	(0.014)	(0.015)
C	0.035	-0.027	-0.173**	-0.233**	-0.017**	-0.014	-0.017
	(0.033)	(0.068)	(0.031)	(0.065)	(0.005)	(0.017)	(0.017)
D	0.183**	0.293**	-0.065**	0.034	-0.001	-0.007	-0.006
	(0.025)	(0.054)	(0.023)	(0.052)	(0.004)	(0.013)	(0.013)
E	0.070*	0.090	-0.081**	-0.061	-0.012**	-0.023	-0.023
	(0.028)	(0.060)	(0.027)	(0.056)	(0.004)	(0.015)	(0.015)
F	0.276**	0.363**	-0.106**	0.012	-0.015**	-0.027*	-0.030*
	(0.025)	(0.055)	(0.026)	(0.051)	(0.004)	(0.013)	(0.013)
G	0.026	0.008	-0.207**	-0.185**	-0.031**	-0.015	-0.014
	(0.035)	(0.073)	(0.034)	(0.061)	(0.005)	(0.019)	(0.019)
H	-0.049	-0.074	-0.086*	-0.034	0.003	-0.001	-0.001
	(0.032)	(0.072)	(0.034)	(0.064)	(0.006)	(0.014)	(0.014)
I	-0.550**	-0.746**	-0.397**	-0.423**	-0.054**	-0.049*	-0.050*
	(0.040)	(0.085)	(0.041)	(0.089)	(0.007)	(0.019)	(0.020)
Mean	2.787	2.962	2.787	2.962	0.491	0.281	0.285
Observations	592153	392094	592153	392094	592153	592153	592153

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Note: Column 1 repeats the specification from Table 2, 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 2 notes for additional specification details. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Appendix Table A8: 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 Months Balanced	(5) 12 Months Balanced
High-Cost Plans	0.120**	0.123**	0.130**	0.108	0.098
	0.038	0.040	0.042	0.087	0.093
Low-Cost Plans	-0.159**	-0.154**	-0.154**	-0.132**	-0.189**
	0.021	0.022	0.023	0.036	0.039
Observations	393570	492483	557304	237339	221562

Standard errors in second row

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Note: Table reports IV estimates of each plan grouping's causal effect on utilization, using a modification to the IV regression in Equation 3. 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 A6, which includes only the first six months post-assignment. See Table A6 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 A9: IV Results Stratifying by Beneficiary Characteristics

	(1) By Sex	(2) By Age	(3) By Baseline Spending
Low-Cost Plans	-0.155**	-0.152**	-0.110**
	(0.026)	(0.024)	(0.024)
Female=1 × Low-Cost Plans=1	-0.012		
	(0.038)		
Above Age 33=1 × Low-Cost Plans=1		-0.015	
		(0.042)	
Baseline Spending=1 × Low-Cost Plans=1			-0.066+
			(0.036)
Observations	393570	393570	393570

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Note: Table reports IV estimates of each plan grouping's causal effect on utilization, using a modification to the IV regression in Equation 3. 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. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A10: Robustness of IV Plan Spending Results to the Inclusion of Controls for Network Breadth and Provider Fixed Effects

	Full Sample		Physician/Hospital Sample			
	(1)	(2)	(3)	(4)	(5)	(6)
High-Cost Plans	0.120**	0.100**	0.096^{+}	0.078	0.075	0.056
	(0.038)	(0.038)	(0.049)	(0.062)	(0.049)	(0.061)
Low-Cost Plans	-0.159**	-0.164**	-0.125**	-0.102**	-0.130**	-0.106**
	(0.021)	(0.021)	(0.027)	(0.028)	(0.027)	(0.028)
Physician/Hospital FEs				X		X
Network Breadth		X			X	X
Observations	393,570	393,570	232,956	232,956	232,956	232,956

Note: Table reports IV estimates of each plan-group's causal effect relative to the omitted, medium-cost plan group, using a modified version of the IV regression in Equation 3. Column 1 reproduces our primary result. Column 2 instruments for provider network breadth using the network breadth of plan of assignment—in addition to instrumenting for plan with plan of assignment (see Appendix Section D.1 for additional details). Columns 3-6 restrict the sample to enrollees we could attribute to a provider (based on the plurality of their healthcare spending). Column 3 presents the results of estimating a modified version of the IV regression in Equation 3 on this subsample of enrollees. Columns 4 and 6 include fixed effects for enrollees' attributed provider. Columns 5 and 6 include instruments for provider network breadth using the network breadth of plan of assignment. The dependent variable is log spending, as in the main specification in Table 2. 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 2 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. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A11: Robustness of IV Plan Spending Results to the Inclusion of Controls for Network Breadth and Provider Fixed Effects

	Full Sample		Physician/Hospital Sample			
	(1)	(2)	(3)	(4)	(5)	(6)
A	-0.096**	-0.019	-0.017	0.006	0.075	0.086+
	(0.035)	(0.039)	(0.048)	(0.047)	(0.051)	(0.051)
B	-0.162**	-0.097**	-0.153**	-0.125*	-0.076	-0.058
	(0.036)	(0.037)	(0.051)	(0.053)	(0.052)	(0.053)
C	-0.157**	-0.172**	-0.155**	-0.102*	-0.170**	-0.115*
	(0.036)	(0.036)	(0.047)	(0.050)	(0.047)	(0.049)
D	0.131**	0.151**	0.110*	0.102	0.135**	0.123+
	(0.041)	(0.041)	(0.052)	(0.064)	(0.052)	(0.063)
E	0.050	0.086**	0.064	0.101*	0.109*	0.140**
	(0.033)	(0.033)	(0.046)	(0.050)	(0.047)	(0.051)
F	-0.024	0.116**	-0.013	-0.017	0.152**	0.126*
	(0.031)	(0.037)	(0.043)	(0.045)	(0.054)	(0.055)
G	-0.123**	-0.055	-0.064	-0.040	0.017	0.031
	(0.034)	(0.036)	(0.047)	(0.049)	(0.051)	(0.053)
H	-0.181**	-0.168**	-0.126*	-0.089+	-0.111*	-0.075
	(0.038)	(0.038)	(0.051)	(0.052)	(0.050)	(0.052)
I	-0.203**	-0.090	-0.201+	-0.138	-0.067	-0.022
	(0.069)	(0.073)	(0.103)	(0.106)	(0.106)	(0.108)
Physician/Hospital FEs				X		X
Network Breadth		X			X	X
Observations	393,570	393,570	232,956	232,956	232,956	232,956
Plan Effect Std. Dev	0.105	0.108	0.097	0.081	0.105	0.087
Plan Effect Range	0.334	0.323	0.311	0.240	0.322	0.255
F-Stat	12.183	14.580	5.648	3.433	7.510	4.676

Note: Table reports IV estimates of each plan's causal effect relative to an omitted plan (X), using a modified version of the IV regression in Equation 3. Column 1 reproduces our primary result. Column 2 instruments for provider network breadth using the network breadth of plan of assignment—in addition to instrumenting for plan with plan of assignment (see Appendix Section D.1 for additional details). Columns 3-6 restrict the sample to enrollees we could attribute to a provider (based on the plurality of their healthcare spending). Column 3 presents the results of estimating a modified version of the IV regression in Equation 3 on this subsample of enrollees. Columns 4 and 6 include fixed effects for enrollees' attributed provider. Columns 5 and 6 include instruments for provider network breadth using the network breadth of plan of assignment. The dependent variable is log spending, as in the main specification in Table 2. 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 2 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. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A12: Robustness of Reduced Form Plan Satisfaction Results to the Inclusion of Controls for Network Breadth and Provider Fixed Effects

	Full Sample		Physician/Hospital Sample			
	(1)	(2)	(3)	(4)	(5)	(6)
High-Cost Plans	0.028**	0.021**	0.046**	0.067**	0.033**	0.056**
	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
Low-Cost Plans	-0.032**	-0.034**	-0.052**	-0.060**	-0.055**	-0.063**
	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
Constant	0.924**	0.868**	0.888**	0.891**	0.792**	0.812**
	(0.002)	(0.005)	(0.002)	(0.003)	(0.008)	(0.009)
Physician/Hospital FEs				X		X
Network Breadth		X			X	X
Observations	65,592	65,592	38,824	35,771	38,824	35,771

Note: Table reports reduced form estimates of each plan-group's causal effect relative to the omitted, medium-cost plan group, using a modified version of a reduced form regression based on Equation 3. Column 1 reproduces our primary result. Column 2 instruments for provider network breadth using the network breadth of plan of assignment—in addition to instrumenting for plan with plan of assignment (see Appendix Section D.1 for additional details). Columns 3-6 restrict the sample to enrollees we could attribute to a provider (based on the plurality of their healthcare spending). Column 3 presents the results of estimating a modified version of the IV regression in Equation 3 on this subsample of enrollees. Columns 4 and 6 include fixed effects for enrollees' attributed provider. Columns 5 and 6 include instruments for provider network breadth using the network breadth of plan of assignment. The dependent variable is willingness-to-stay (i.e., consumer satisfaction), as in the main specification in Table 2. 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 2 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. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A13: Robustness of Reduced Form Plan Satisfaction Results to the Inclusion of Controls for Network Breadth and Provider Fixed Effects

	Full Sample		Physician/Hospital Sample			
	(1)	(2)	(3)	(4)	(5)	(6)
A	-0.041**	-0.024**	-0.067**	-0.073**	-0.039**	-0.046**
	(0.004)	(0.004)	(0.007)	(0.007)	(0.008)	(0.008)
B	-0.068**	-0.055**	-0.103**	-0.104**	-0.081**	-0.083**
	(0.004)	(0.004)	(0.006)	(0.007)	(0.006)	(0.007)
C	-0.015**	-0.018**	-0.018*	-0.011	-0.023**	-0.015+
	(0.005)	(0.005)	(0.007)	(0.008)	(0.007)	(0.008)
D	0.019**	0.023**	0.031**	0.059**	0.038**	0.065**
	(0.005)	(0.005)	(0.007)	(0.008)	(0.007)	(0.008)
E	-0.028**	-0.020**	-0.043**	-0.043**	-0.030**	-0.031**
	(0.004)	(0.004)	(0.007)	(0.008)	(0.007)	(0.008)
F	-0.006	0.025**	-0.011*	0.008	0.039**	0.056**
	(0.004)	(0.005)	(0.005)	(0.007)	(0.008)	(0.009)
G	-0.056**	-0.041**	-0.090**	-0.084**	-0.064**	-0.059**
	(0.004)	(0.004)	(0.006)	(0.007)	(0.007)	(0.008)
H	-0.030**	-0.027**	-0.051**	-0.057**	-0.047**	-0.053**
	(0.005)	(0.005)	(0.008)	(0.009)	(0.008)	(0.009)
I	-0.047**	-0.023**	-0.082**	-0.083**	-0.044**	-0.047**
	(0.008)	(0.008)	(0.014)	(0.015)	(0.015)	(0.016)
Constant	0.935**	0.852**	0.905**	0.902**	0.770**	0.772**
	(0.002)	(0.010)	(0.004)	(0.004)	(0.015)	(0.017)
Physician/Hospital FEs				X		X
Network Breadth		X			X	X
Observations	65,592	65,592	38,824	35,771	38,824	35,771
Plan Effect Std. Dev	0.026	0.024	0.041	0.049	0.038	0.047
Plan Effect Range	0.087	0.080	0.134	0.164	0.120	0.149
F-Stat	62.125	60.858	66.465	72.221	64.393	67.394

Note: Table reports reduced form estimates of each plan's causal effect relative to an omitted plan (X), using a modified version of a reduced form regression based on Equation 3. Column 1 reproduces our primary result. Column 2 instruments for provider network breadth using the network breadth of plan of assignment—in addition to instrumenting for plan with plan of assignment (see Appendix Section D.1 for additional details). Columns 3-6 restrict the sample to enrollees we could attribute to a provider (based on the plurality of their healthcare spending). Column 3 presents the results of estimating a modified version of the IV regression in Equation 3 on this subsample of enrollees. Columns 4 and 6 include fixed effects for enrollees' attributed provider. Columns 5 and 6 include instruments for provider network breadth using the network breadth of plan of assignment. The dependent variable is willingness-to-stay (i.e., consumer satisfaction), as in the main specification in Table 2. 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 2 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. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A14: 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	0.005*	-0.002	-0.019**	0.008	0.034**	-0.002	0.010*
	(0.002)	(0.003)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
Low-Cost Plans	-0.002**	-0.005**	-0.008**	-0.013**	-0.010**	-0.010**	-0.011**
	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Mean	0.022	0.055	0.094	0.210	0.082	0.083	0.195
Observations	393570	393570	393570	393570	393570	393570	393570

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Note: Table reports IV regression results for category or place of service, using a modification to the IV regression in Equation 3. 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 4 in Section 3.1. 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 2 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 A15: Main IV Results for High-Value Care Measures

	(1) Primary Care	(2) HbA1c Testing	(3) Breast Cancer Screening	(4) Cervical Cancer Screening	(5) Chlamydia Screening
High-Cost Plans	0.029**	-0.000	0.000	0.002*	0.000
	(0.005)	(0.001)	(0.000)	(0.001)	(0.001)
Low-Cost Plans	-0.001	-0.001**	0.000	-0.001*	-0.001**
	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Mean	0.103	0.005	0.001	0.007	0.007
Observations	393570	393570	393570	393570	393570

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Note: Table reports IV regression results for use of "high-value care," using a modification to the IV regression in Equation 3. 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 A14. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A16: Main IV Results for Utilization of Select Drug Categories

	(1) Diabetes	(2) Statins	(3) Anti-Depressants	(4) Anti-Psychotics	(5) Anti-Hypertension	(6) Anti-Stroke	(7) Asthma	(8) Contra-ceptives
High-Cost Plans	0.001	0.005*	0.002	-0.002	0.003	0.001	-0.000	-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)
Mean	0.022	0.024	0.034	0.038	0.033	0.003	0.016	0.008
Observations	393570	393570	393570	393570	393570	393570	393570	393570

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Note: Table reports IV regression results for prescription drug fills, using a modification to the IV regression in Equation 3. 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 A14. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A17: Main IV Results for Low-Value Care Measures

	(1) Abdomen CT	(2) Thorax CT	(3) Imaging and Lab	(4) Head Imaging for Uncomp. HA	(5) Avoidable Hospitalizations
High-Cost Plans	0.000	0.000	0.006	0.000	0.001
	(0.000)	(0.000)	(0.005)	(0.000)	(0.001)
Low-Cost Plans	0.000	0.000	-0.009**	-0.000	0.001+
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)
Mean	0.000	0.000	0.144	0.002	0.005
Observations	393570	393570	393570	393570	393570

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Note: Table reports IV regression results for use of "low-value care," using a modification to the IV regression in Equation 3. 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 A14. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A18: 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	-0.293**	-0.296**	-0.312**	-0.316**
	(0.038)	(0.031)	(0.030)	(0.027)
B	-0.528**	-0.425**	-0.526**	-0.513**
	(0.043)	(0.030)	(0.031)	(0.030)
C	0.027	-0.173**	-0.142**	-0.231**
	(0.036)	(0.031)	(0.032)	(0.030)
D	0.200**	-0.065**	0.393**	0.295**
	(0.026)	(0.023)	(0.022)	(0.020)
E	0.069*	-0.081**	0.021	-0.034
	(0.033)	(0.027)	(0.026)	(0.025)
F	0.281**	-0.106**	0.293**	0.094**
	(0.032)	(0.026)	(0.023)	(0.020)
G	0.016	-0.207**	-0.280**	-0.405**
	(0.045)	(0.034)	(0.029)	(0.024)
H	-0.057	-0.086*	-0.227**	-0.265**
	(0.041)	(0.034)	(0.034)	(0.035)
I	-0.542**	-0.397**	-0.460**	-0.333**
	(0.046)	(0.041)	(0.037)	(0.039)
Mean (dollars)	$460	$460	$480	$480
County x Year x Month FEs	X	X	X	X
Person-Level Controls		X		X
Obs: Enrollee X Months	592153	592153	985723	985723

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 2. 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 2 for additional details on the specifications. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Appendix Table A19: 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	0.154*	0.167*	488.475**	0.123**	0.200**	528.744**
	(0.069)	(0.075)	(163.363)	(0.036)	(0.069)	(155.450)
Low-Cost Plans	-0.317**	-0.341**	-203.746**	-0.133**	-0.300**	-174.978**
	(0.031)	(0.034)	(63.218)	(0.019)	(0.031)	(59.180)
Mean	4.41	4.84	2732.90	2.09	4.40	2602.14
Observations	65595	65595	65595	65595	65595	65595

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Note: Table reports IV estimates of each plan grouping's causal effect on utilization, using a modification to the IV regression in Equation 3. 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. (4) in Section 3.1. The difference here compared to Table A6 is that spending and utilization outcomes are totalled over the full six-month enrollment spell. Specifications otherwise follow Tables A5 and A6. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.

Appendix Table A20: No Differential Attrition Out of Medicaid Program Across Plan of Assignment

	(1) 6 Months	(2) 12 Months	(3) 18 Months	(4) 24 Months
A	0.003	0.008	0.003	-0.000
	(0.004)	(0.010)	(0.010)	(0.009)
B	-0.005	-0.015	0.011	0.043**
	(0.005)	(0.011)	(0.011)	(0.011)
C	-0.003	-0.013	-0.022+	-0.017+
	(0.005)	(0.013)	(0.011)	(0.010)
D	-0.003	0.012	0.011	0.008
	(0.006)	(0.013)	(0.013)	(0.012)
E	-0.007	-0.014+	-0.026*	-0.065**
	(0.004)	(0.008)	(0.011)	(0.010)
F	-0.002	-0.006	-0.009	-0.003
	(0.004)	(0.011)	(0.010)	(0.008)
G	-0.005	-0.002	0.003	0.004
	(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	-0.007	-0.017	-0.013
	(0.007)	(0.014)	(0.012)	(0.011)
Mean	.955	.344	.269	.197
Observations	33902	33902	33902	33902

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

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 February 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. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$ $p < 0.01$.