



Published in final edited form as:

Am Econ J Appl Econ. 2023 July ; 15(3): 341–379. doi:10.1257/app.20210843.

What Difference Does a Health Plan Make? Evidence from Random Plan Assignment in Medicaid

Michael Geruso[†], Timothy J. Layton[‡], Jacob Wallace[§]

[†]UT-Austin and NBER.

[‡]Harvard University and NBER.

[§]Yale School of Public Health.

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 reduces spending by at least 25%—primarily through quantity reductions—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.

1 Introduction

Private managed care plans are the dominant form of healthcare delivery in the United States, including in publicly funded programs (Gruber, 2017). Although there is growing evidence on the roles of hospitals (Doyle et al., 2015; Hull, 2020), nursing homes (Einav, Finkelstein and Mahoney, 2022), and individual physicians (Kwok, 2019) in determining the allocation of healthcare goods and services, evidence on the impact of health plans is less complete. Several studies have documented the effects of demand-side cost sharing tools such as deductibles and copays.¹ But a modern health plan consists of much more than just a schedule of consumer-facing prices, and relatively little is known about the extent to which plans can and do use other tools to influence healthcare consumption, clinical quality, and satisfaction. There is especially little evidence regarding such plan effects in the largest health insurance program in the United States—Medicaid. In addition to being an important program in its own right, Medicaid is an ideal setting to study how plans use managed care tools to influence the allocation of healthcare because cost-sharing is generally prohibited, so beneficiaries choose between managed care plans differentiated only in their supply-side features (provider networks, utilization management rules, etc.).²

mike.geruso@austin.utexas.edu .

¹For 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, drawing on evidence from Medicaid. In the first part, we ask whether private managed care plans can substantially 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 allocate beneficiaries to plans that efficiently constrain healthcare spending.

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, 2018a,b); and teacher and school effects (Chetty, Friedman and Rockoff, 2014a,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 25% less in healthcare spending than if the same individual enrolled in the highest-spending plan in the market.³ This finding is, in itself, a striking new fact. To put this result in context, a 25% difference in total mean spending was close to 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.⁴

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

²In 2017, 84% of Medicaid beneficiaries (62 million) were enrolled in a private managed care plan (Kaiser Family Foundation, 2019a). In the same year, almost \$500 billion of the \$1.3 trillion spent on public health insurance programs went to private insurers (Kaiser Family Foundation, 2017, 2019a). By 2018, Medicaid spending was nearly 30% of overall state spending (Figure 1).

³This 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).

⁴This 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.

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 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.⁵ 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).⁶ 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 and that lower-spending plans (especially the three for-profits) are less likely to attract enrollees making active choices, including noncomplying enrollees switching away from their plan of assignment. This suggests a real trade-off between spending and beneficiary satisfaction, a

⁵These 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.

⁶We 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).

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, 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).⁷ 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. 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.

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

⁷Due 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).

South Carolina by Garthwaite and Notowidigdo (2019) and in Louisiana by Wallace et al. (2022), and health plans serving the non-elderly, non-Medicaid population by Handel et al. (2019).⁸ This paper is also closely related to Wallace (2020), which uses the same random assignment in New York we use here, combined with within-plan geographic variation, to study how narrow provider networks affect beneficiary outcomes.

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—with effects larger than what a high deductible has been shown to accomplish—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.”⁹ 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.

⁸Wallace et al. (2022) uses the same type of identifying variation (the auto-assignment of Medicaid enrollees to plans), but its setting is Louisiana Medicaid, where the study population is largely healthy children and pregnant women, rather than New York Medicaid, where our study population is adults. The objective in Wallace et al. (2022) is also different: to estimate whether risk-adjusted measures of plan performance adequately adjust for the variation across plans that arises because of differences in patient characteristics (residual confounding).

⁹In 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.

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.¹⁰ 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.

Plans' incentives for cost control and patient satisfaction were determined by the combination of the contract structure and the institutional feature that all plans would receive some enrollees through the auto-assignment process regardless of patient satisfaction or quality (above the state's minimum threshold). Plans received a monthly capitation payment for each individual enrolled in the plan in a given month. In the beginning of the study period, these payments were plan-specific and were based on each plan's spending patterns in prior years. By the end of the period, payments were set at the market level (i.e., made uniform across plans) and risk-adjusted according to each enrollee's clinical conditions. Thus, throughout the study period, plans were the residual claimants on any healthcare spending reductions in a given plan-year, though the incentive to constrain spending was stronger later in the study period because of the dynamic incentives involved in rate-setting.¹¹

In terms of attracting and retaining enrollees, plans may have had asymmetric strategies enabled by the presence of auto-assignment and enrollment inertia in this context. Below, we show that the pattern of enrollment and spending across plans would be consistent with some plans pursuing a high-margin, low-volume strategy and others pursuing a low-margin, high-volume strategy.¹²

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 (New York State Department of Health, 2008–2012). 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

¹⁰See Appendix A for additional detail.

¹¹When payment rates were based on each plan's lagged spending, each plan was the residual claimant on the static savings from the current year, but payments in future years would be lower as a result of any current savings. After the transition to market-level rate setting, each plan was the residual claimant on the static savings from the current year relative to the market-level rate, and future payments would only be affected to the extent that the plan's own reduced spending lowered market-level average spending (a smaller dynamic disincentive to constrain spending).

¹²The existence of such an equilibrium in the presence of inattentive consumers is a key finding of models of price dispersion (Salop and Stiglitz, 1982). Passive auto-assignees, who take-up these managed care products via administrative assignment, are perhaps the ultimate inattentive consumers.

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). For example, as low-value care measures we assess the likelihood an enrollee uses the emergency department for avoidable reasons (Medi-Cal Managed Care Division, 2012) and the hospital all-cause readmission rate (National

Committee for Quality Assurance, NCQA) (Washington Health Alliance, 2015). We use General Equivalence Mappings from ICD-10 CM diagnosis codes to ICD-9 codes to assign low value status to procedures in the data (CMS, 2018). We also use the IBM Micromedex RED BOOK to classify drugs into therapeutic classes (IBM, 2020).

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 5.2, 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.¹³ 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.

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

¹³The sample size of auto-assignees is not identical across plans for several reasons. First, plans qualify to receive auto-assignees based on a 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 county \times month \times year of enrollment fixed effects, as within a county \times month \times year, assignment is purely random. We link enrollees to a county and a neighbourhood using their zip code of residence (United States Census Bureau, 2010) (United Hospital Fund, 2004). In principle, plans could reach capacity constraints and so be unwilling to accept new enrollees and auto-assignees. In practice, there is no evidence this ever occurred during our study period. Figure A1 shows that each month, each eligible plan received a $\frac{1}{N}$ share of enrollees, where N is the number of eligible plans.

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.¹⁴ 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.¹⁵ Third, we restrict to beneficiaries with at least six months of post-assignment enrollment in Medicaid to allow us to observe plan effects on spending, utilization, and quality outcomes.

In primary analyses we restrict attention to the initial six months post-assignment. Enrollment is high and stable until six months and then drops off precipitously (see Appendix Figure A2). 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, further detailed in Appendix Table A1, leave us with 65,595 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.¹⁶ In some analyses, we compare outcomes between auto-assigned beneficiaries and those that made active plan choices. To do so, we construct a 10% random sample of “active-choosers” (for computational feasibility), imposing the same basic sample restrictions we used for the auto-assigned beneficiaries. However, when testing for balance across plans on predetermined characteristics we construct a random subsample of this 10% active-chooser sample that is equal size to the auto-assignee sample to equalize statistical power across the two groups.

¹⁴After 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 effects.

¹⁵Auto 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.

¹⁶For summary statistics separated by the pre- and post-assignment periods, please refer to Appendix Table A2.

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) or beneficiary-time (it) level. We follow Finkelstein, Gentzkow and Williams (2016) in modeling a data generating process for healthcare spending in which log spending (Y) is determined by a plan component (γ), a person-level component (ξ), time-varying observables (X), and a mean zero shock (ϵ).¹⁷ To recover plan effects, γ_j , we estimate regressions of the following form, combining the individual component, ξ , and the error, ϵ , into a compound error term μ :

$$Y_{it} = \rho + \psi_{c(i)} + vX_{it} + \sum_{j=1}^9 \gamma_j Plan_j_{it} + \mu_{it}. \quad (1)$$

In these regressions, an observation is a beneficiary-month.¹⁸ The subscript t denotes the month \times year of the observation. The subscript $c(i)$ denotes an assignment cohort of beneficiary i —fixed for each individual—defined by the county \times month \times year in which the beneficiary was originally assigned to a managed care plan. 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), denoted $Plan_j_{it}$ and equal to 1 if beneficiary i is enrolled in plan j in month t and zero otherwise. Assignment cohort fixed effects $\psi_{c(i)}$ (e.g., an indicator for being assigned in June 2011 in the Bronx) are included in all specifications, as this is the unit of randomization. The X vector of individual controls includes indicators for: sex, race (5 groups), deciles of spending in FFS prior to MMC enrollment, and each individual year of age (18 to 64).

To address the endogeneity of beneficiaries sorting across plans—correlation between plan choice and μ_{it} ($\equiv \xi_i + \epsilon_{it}$)—we exploit random assignment, which is cross-sectional within an assignment cohort. We restrict to individuals who were randomly auto-assigned to plans and instrument for plan *enrollment* indicators with plan *assignment* indicators, denoted $Assigned_j_i$ and equal to 1 if the beneficiary was assigned to plan j and 0 otherwise. There are ten plans that receive auto-assigned enrollees during our time period, requiring nine first-stage regressions (with plan 10 omitted):

$$\begin{aligned} Plan_1_{it} &= \alpha_1 + \phi_{1c(i)} + \delta_1 X_{it} + \sum_{j=1}^9 \lambda_{1j} Assigned_j_i + \eta_{1,it} \\ &\vdots \\ Plan_9_{it} &= \alpha_9 + \phi_{9c(i)} + \delta_9 X_{it} + \sum_{j=1}^9 \lambda_{9j} Assigned_j_i + \eta_{9,it}. \end{aligned} \quad (2)$$

¹⁷FGW 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 $Y_{it} = vX_{it} + \gamma_j + \xi_i + \epsilon_{it}$.

¹⁸In 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.

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 ($\widehat{\mathbf{Plan}}_{it}$) from the first-stage regressions:

$$Y_{it} = \rho + \psi_{c(i)} + \nu X_{it} + \sum_{j=1}^9 \gamma_j \widehat{\mathbf{Plan}}_{j_{it}} + \mu_{it}. \quad (3)$$

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:

$$Y_{it} = \rho + \psi_{c(i)} + \nu X_{it} + \gamma_{\text{Low}} \widehat{\mathbf{LowPlan}}_{it} + \gamma_{\text{High}} \widehat{\mathbf{HighPlan}}_{it} + \mu_{it} \quad (4)$$

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 2.¹⁹

3.2 First-Stage and Instrument Validity

Panel (b) of Figure 2 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 effects recovered by IV are unlikely to differ much from average treatment effects for the full auto-assignee sample. Table A3 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. Panel (a) of Figure 2 presents a series of balance 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. These characteristics are measured at the enrollee level rather than the enrollee-month level (collapsed to pre-assignment means for time-varying outcomes like

¹⁹Specifically, the first stage for enrollment in the “low” group is $\text{LowPlan}_{it} = \alpha_{\text{low}} + \phi_{\text{low},c(i)} + \delta_{\text{low}} X_{it} + \lambda_{\text{low,low}} \text{Assigned_low}_{it} + \lambda_{\text{low,high}} \text{Assigned_low}_{it} + \eta_{\text{low},it}$

monthly spending prior to assignment). 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 panel (a) Figure 2 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 panel (a) of Figure 2, 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.²⁰ Successful random assignment would 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

²⁰Tabular versions of these results are in Table A4.

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 5, discussed below).²¹

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 City. This limits the external validity of our estimates due to the reliance on auto-assignees, who are a non-random sample of the Medicaid population, and our focus on New York City, which is an urban market that differs from other parts of the country.

Auto-assignment only occurs if enrollees don't select a plan.²² It is therefore useful to understand whether auto-assignees differ on observables from active choosers. Table A5 shows that auto-assignees differ somewhat from active-choosers on observables, 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 that use typical levels of care, rather than enrollees that are not actively engaged with the health care system. To maximize the 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. However, it is important to point out that the auto-assignee population may also differ from the average Medicaid enrollee on unobserved characteristics. For example, if it is easier to restrict access to care for auto-assignees—because, for example, they are selected on passivity—then the estimated range of spending effects in this paper may overstate the scope for plans to generate spending differences for the typical Medicaid enrollee.

Beyond the issues of generalizing from the auto-assignees to the full *New York* Medicaid population, it is useful to compare New York Medicaid to Medicaid programs elsewhere. Appendix Table A7 compares the characteristics of state Medicaid programs, including Medicaid spending per enrollee (Centers for Medicaid and Medicare Services, 2011), the fraction of Medicaid beneficiaries in managed care plans (United States Census Bureau, 2019), the number of plans serving in the state (Kaiser Family Foundation, 2019b), the share of these plans that are for-profit, and the share of Medicaid enrollees who are auto-assigned.

²¹Tabular versions of plan-level attrition results present a similar story (Table A6).

²²One potential pathway into the Medicaid program for auto-assignees may be via presumptive eligibility, a process by which health care professionals or hospitals can enroll uninsured patients in Medicaid if they deem them eligible for Medicaid. Given that our study sample is composed of individuals that do not select their plans, it is possible that a disproportionate share of these enrollees may have entered the program via presumptive eligibility.

The table shows that the Medicaid program in New York is typical in some ways: it contracts with a variety of plans and plan types, including national for-profits and various not-for-profits, and it uses auto-assignment. New York is atypical in other ways: A larger share of the population resides in urban areas (Iowa State University, Iowa Community Indicators Program, 2022), Medicaid managed care penetration is higher than other states, and a relatively small share of enrollees are auto-assigned. Also, the fact that New York has an unusually large number of plans competing in the same market suggests that other states may not feature as much between-plan variation as we find in our study context. Given these considerations, the large range of plan effects we document below could be informative of the *potential* for plans to have large impacts on spending even without the ability to set cost-sharing, but may not be predictive of the range of plan effects in any particular Medicaid managed care market.

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. Six plans (A, B, C, G, H, I) spent significantly less than the omitted plan (X), two plans (E and F) had spending levels similar to the omitted plan, and one plan (D) had significantly higher spending. Interestingly, the three lowest-spending plans (B, H, I) are the three for-profit plans in our setting. Comparing the highest-spending plan, D (+13.1% relative to the omitted plan), to the lowest-spending plan, I (−20.3% relative to the omitted plan), yields an overall range of about 33 percentage points. With an inference on winners correction (Andrews, Kitagawa and McCloskey, 2019), this range shrinks to 25 percentage points.²³ 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 a similar range in plan effects when the outcome is parameterized as the inverse hyperbolic sine of spending or Winsorized spending levels (Table A8), when we use a Poisson regression that places less emphasis on zeros relative to the log specification (Table A9), and when we aggregate spending over the entire six-month enrollment spell, rather than analyzing monthly outcomes (Table A10). Each approach is affected differently by the presence of zero-spending months at the person \times month level, but all are broadly consistent: In some specifications, some of the plan effects of the “interior” plans attenuate, but the range between the highest and lowest spending plans is quite robust, always about

²³We adapt the method of Andrews, Kitagawa and McCloskey (2019), applying their inference correction to determine (separately) adjustments to the highest and lowest coefficients. Below, when using plan-specific coefficients as observations in a regression, we use an empirical Bayes procedure to shrink estimated plan effects. And when reporting on the standard deviation of plan effects, we report an adjusted (shrunk) standard deviation that likewise corrects for error in the estimates.

the same range as in Figure 3. Further, the “any use” results in Figure 3 show clear evidence of significant variation in extensive margin plan effects on utilization, an outcome that is unaffected by outliers in the healthcare spending distribution and that is not subject to the issues caused by zeros in log or inverse hyperbolic sine specifications.

The similarity of plan effects on total spending and plan effects on an indicator for any utilization suggests that quantity differences may be more important than negotiated price differences in this context. To illustrate how much of the spending differences can be accounted for by prices, panel (a) of Figure 3 also reports coefficients from a version of the IV regression in which all claims have been repriced as if every plan transacted with providers at a common set of prices. Repricing has almost no effect on our estimates of plan spending coefficients, indicating that price differences cannot account for the spending differences we observe.²⁴

The range of these estimates is large.²⁵ 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). This fact remains true even after correcting for the noisiness of the estimates: In Table 2, the standard deviation of estimated plan effects shrinks only a little when one adjusts for the standard error of the estimates (.102 vs .112).²⁶ 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 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 A11.²⁷

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

²⁴See Appendix D.2 for additional, more granular price comparisons between the higher and lower spending plans.

²⁵External 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.

²⁶The adjusted standard deviation of the plan effects is calculated as $\hat{\sigma} = \sqrt{J^{-1} \sum_{j=1}^J [(\hat{\gamma}_j - \bar{\gamma})^2 - SE_j^2]}$.

²⁷The reverse exercise—reweighting the auto-assignees to resemble the active choosers in the plan effects regression—is reported in Table A8.

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 Effects over Enrollment Spells

How do these effects unfold over time? 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- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high-spending (Plan D) groups based on the IV spending effects as described above.²⁸ 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.²⁹ 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. 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.³⁰ 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.

Figure 4 shows that the effects begin immediately upon enrollment and then are generally stable over the entire enrollment spell. In the baseline sample (panel (a)), 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 A12 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

²⁸The group-level spending differences between low and medium and between medium and high plans are 16 and 11 percent, respectively (Table A13).

²⁹In particular, $\log(\text{Spending} + 1)_{it}^e = \rho^\tau + \psi_{e(t)}^\tau + v^\tau X_i + \gamma_{\text{Low}}^\tau \widehat{\text{LowPlan}}_i + \lambda_{\text{High}}^\tau \widehat{\text{HighPlan}}_i + \mu_i^\tau$, which is estimated separately for each $\tau \in -2, -1, 0, 1, 2, 3, 4, 5$.

³⁰This can be seen in Figure A2. 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.

on spending are remarkably consistent across these samples. Overall, Figure 4 and Table A12 show that spending effects remain large throughout the post-assignment months.³¹

4.3 Heterogeneity

In Figure A3 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 A3 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 A14 presents pooled regression results corresponding to each of the event studies in Figure A3, 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 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.

5.1 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

³¹The figure also rules out that our findings are driven by temporary disruption effects: Some disruption is likely to occur in any new plan transition. However, 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 common transition effect would be differenced-out in the low versus medium plan comparison.

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 5, 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 5 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 have limited effects on the utilization of most of these high value drugs, with some suggestive evidence that they lower utilization for some. 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 5 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 5.3.

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 5 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.

5.2 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.” We measure willingness-to-stay at the individual level with a indicator variable equal to 1 if the individual is enrolled in their assigned plan 3 (or 6) months post-assignment and 0 otherwise. 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.

Plan effects on willingness-to-stay are reported in the last two columns of Table 2. These coefficient estimates range from -0.068 (Plan B) to 0.019 (plan D), implying a range of differences in willingness-to-stay of 8.7 percentage points. Relative to the baseline rate of remaining in the assigned plan among those assigned to the omitted plan (90.6%), this represents an almost 10% difference, which we interpret as economically meaningful, especially given the likely high levels of inertia and inattention. Put another way, in the omitted plan, 9.4% of assigned beneficiaries left the plan by 6 months post-assignment. In Plan B, 16.2% ($= 9.4\% + 6.8\%$) of assigned beneficiaries left the plan by 6 months post-assigned, about a 70% difference in this exit rate.

The enrollee satisfaction panel of Figure 5 shows that people are less likely to stay in lower-spending plans. This figure also shows how willingness-to-stay evolves over time, measuring willingness-to-stay at 3 and 6 months post-assignment. Overall, willingness-to-stay is lower in lower-spending 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 5.3.

To give a finer view of these results, in Figure 6 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 they 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. In Section 6.1 we present additional evidence on the relationship between plan spending and enrollment flows, including among active-chooser beneficiaries.

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 5 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.³² This suggests that the tools used by low-spending plans to constrain costs could have negative consequences for beneficiary health.

5.3 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

³²The point estimate for the effect of assignment to a lower-spending plan on avoidable emergency department utilization was also positive, but imprecise and statistically insignificant. We did not find any significant relationship between assignment to a lower-spending plan and hospital readmission rates. See Figure 5. Interpreting the readmission result is complicated by our finding that an enrollee's plan of assignment impacts the probability of an initial hospital admission, on which readmission is conditioned.

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).³³ 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.³⁴

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 styles (Glied, 2000). To assess whether efficient steering could explain

³³Appendix Tables A15–A18 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.

³⁴See 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.

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 A15–A16).³⁵ Further, plans are not paying providers differently in a way that could account for our findings: Figure 3 showed plan differences in real resource use, not differences in negotiated provider rates, were responsible for plan spending effects. 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,³⁶ 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.

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

³⁵Besides providers' role in spending, patient-provider satisfaction, trust, and communication may play an important role in patient satisfaction, though we cannot directly assess these features in our data, other than controlling for provider fixed effects.

³⁶This 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.

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 7 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 4.5 percentage point higher market share among enrollees making active choices (R^2 of 0.74). 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 7 results show that demand seems to primarily be tied to the ability to use care, and thus to higher levels of healthcare spending.

Figure 8 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 8 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 8. The lowest spending plan, I, attracts the lowest (nearly zero) enrollment shares by any of the flow measures in Figure 8. In contrast, among

auto-assignees, plans receive equal *initial* shares of enrollment flows by the randomization design.

Of course, just as there is evidence of significant choice imperfections from a variety of studies of health plan choice, there are likely significant frictions, inertia, and mistakes in the choices represented in 7 and 8. Further, depending on what one views as the objective in Medicaid—focused narrowly on avoiding certain preventable adverse health outcomes or more broadly inclusive of the type of amenities that healthcare consumers in other contexts value, such as convenient access to care—one may place different weight on the importance of enrollee satisfaction.³⁷ Here we merely note that satisfaction, as revealed by plan switching, is strongly correlated with plan spending.

Finally, we note that the for-profits (B, H, and I) are among the least preferred plans in Figure 8: In panel (b) (plans chosen by auto-assignees switching out of their assigned plans) they are the strictly less-preferred than all others. And by the other measures of satisfaction in the figure, the for-profits are always among the least preferred.

6.2 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 reallocation that removed the top 4 highest-spending plans in the market and reassigned 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 (9.7% or \$637.56 per active-choosing beneficiary per year).³⁸ Thus, procurement matters for the cost of managed care.

³⁷A focus on enrollee satisfaction would be consistent with the explicit goals of the New York State Medicaid Managed Care program to assess “enrollee satisfaction with care,” which is a performance measure by which plans contracting with the state are evaluated. For additional details please refer to: https://www.health.ny.gov/health_care/managed_care/docs/quality_strategy.pdf.

³⁸The 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. In FY2010, New York State spent \$14.4 billion on Medicaid, $(\$1.4 / \$14.4) = 9.7\%$ (Office of the New York State Comptroller (2021)).

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 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. Potential changes in avoidable emergency department use, hospital readmissions, and other undesirable outcomes are also plausible, though more difficult to measure. Figure 6 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.³⁹ This suggests that states might be able to achieve spending reductions by simply selecting plans with the largest negative OLS risk-adjusted spending effects.⁴⁰

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 over 25 percentage points—a range that exceeds the difference between exposing consumers to a high versus zero deductible (Brot-Goldberg et al., 2017). 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. 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 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 quantity reductions caused by lower-spending plans are

³⁹The 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.

⁴⁰Further, 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.

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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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) and the National Institute on Aging (Grant no. P01AG032952). 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.

References

- Abaluck Jason, Bravo Mauricio Caceres, Hull Peter, and Starc Amanda. 2021. “Mortality effects and choice across private health insurance plans.” *The quarterly journal of economics*, 136(3): 1557–1610. [PubMed: 34475592]
- Abaluck Jason, Caceres Bravo Mauricio M., Hull Peter, and Starc Amanda. 2020. “Mortality Effects and Choice Across Private Health Insurance Plans.” National Bureau of Economic Research Working Paper 27578.
- Andrews Isaiah, Kitagawa Toru, and McCloskey Adam. 2019. “Inference on winners.” National Bureau of Economic Research.

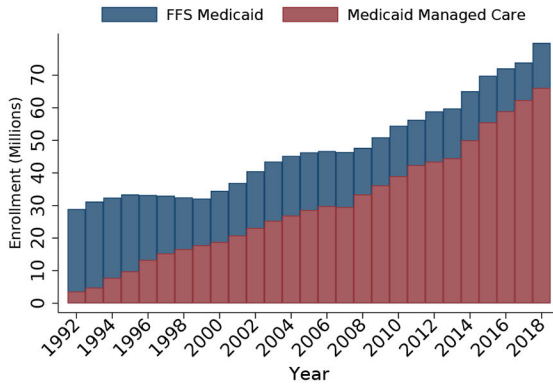
- Angrist Joshua D, Hull Peter D, Pathak Parag A, and Walters Christopher R. 2017. “Leveraging lotteries for school value-added: Testing and estimation.” *The Quarterly Journal of Economics*, 132(2): 871–919.
- Angrist Joshua, Hull Peter, Pathak Parag, and Walters Christopher. 2016. “Interpreting tests of school VAM validity.” *American Economic Review*, 106(5): 388–92.
- Aron-Dine Aviva, Einav Liran, and Finkelstein Amy. 2013. “The RAND Health Insurance Experiment, Three Decades Later.” *Journal of Economic Perspectives*, 27(1): 197–222. [PubMed: 24610973]
- Brot-Goldberg Zarek C, Chandra Amitabh, Handel Benjamin R, and Kolstad Jonathan T. 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, Dodd Allison Hedley, et al. 2015. “Assessing the usability of encounter data for enrollees in comprehensive managed care 2010–2011.” *Mathematica Policy Research*.
- Cabral Marika, Geruso Michael, and Mahoney Neale. 2018. “Do larger health insurance subsidies benefit patients or producers? Evidence from Medicare Advantage.” *American Economic Review*, 108(8): 2048–87. [PubMed: 30091862]
- Centers for Medicaid and Medicare Services. 2011. “Medicaid Managed Care Enrollment Report.” <https://www.kff.org/wp-content/uploads/sites/2/2013/12/2011-medicare-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-medicare-mc-enrollment-report.pdf>.
- Centers for Medicaid and Medicare Services and Mathematica Policy Research. 2016. “Medicaid Managed Care Enrollment and Program Characteristics, 2016.” <https://www.mathematica.org/publications/medicaid-managed-care-enrollment-and-program-characteristics-2016>.
- Centers for Medicaid and Medicare Services and Mathematica Policy Research. 2018. “Medicaid Managed Care Enrollment and Program Characteristics, 2018.” <https://www.medicare.gov/medicaid/managed-care/downloads/2018-medicare-managed-care-enrollment-report-updated.pdf>.
- Chandra Amitabh, Finkelstein Amy, Sacarny Adam, and Syverson Chad. 2016. “Health care exceptionalism? Performance and allocation in the US health care sector.” *American Economic Review*, 106(8): 2110–44. [PubMed: 27784907]
- Charlesworth Christina J, Meath Thomas HA, Schwartz Aaron L, and McConnell K John. 2016. “Comparison of low-value care in Medicaid vs commercially insured populations.” *JAMA internal medicine*, 176(7): 998–1004. [PubMed: 27244044]
- Chernew Michael, Schwartz J Sanford, and Fendrick A Mark. 2015. “Reconciling prevention and value in the health care system.” *Health Affairs*. Blog Post
- Chetty Raj, and Hendren Nathaniel. 2018a. “The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects.” *The Quarterly Journal of Economics*, 133(3): 1107–1162.
- Chetty Raj, and Hendren Nathaniel. 2018b. “The impacts of neighborhoods on intergenerational mobility II: County-level estimates.” *The Quarterly Journal of Economics*, 133(3): 1163–1228.
- Chetty Raj, Friedman John N, and Rockoff Jonah E. 2014a. “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates.” *American Economic Review*, 104(9): 2593–2632.
- Chetty Raj, Friedman John N, and Rockoff Jonah E. 2014b. “Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood.” *American economic review*, 104(9): 2633–79.
- Chetty Raj, Hendren Nathaniel, and Katz Lawrence F. 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. [PubMed: 29546974]
- CMS. 2018. “2018 ICD-10 CM and GEMs.” <https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs>.
- Cooper Zack, Craig Stuart V, Gaynor Martin, and Van Reenan John. 2019. “The Price Ain’t Right? Hospital Prices and Health Spending on the Privately Insured.” *Quarterly Journal of Economics*, 134(1): 51–107. [PubMed: 32981974]
- Cuddy Emily, and Currie Janet. 2020. “Rules vs. discretion: Treatment of mental illness in us adolescents” *National Bureau of Economic Research*.

- Curto Vilsa, Einav Liran, Finkelstein Amy, Levin Jonathan, and Bhattacharya Jay. 2017. "Healthcare Spending and Utilization in Public and Private Medicare." National Bureau of Economic Research Working Paper 23090.
- Curto Vilsa, Einav Liran, Levin Jonathan, and Bhattacharya Jay. 2021. "Can health insurance competition work? evidence from medicare advantage." *Journal of Political Economy*, 129(2): 570–606.
- Cutler David M, McClellan Mark, and Newhouse Joseph P. 2000. "How does managed care do it?" *The Rand journal of economics*, 526–548. [PubMed: 11503704]
- Doyle Joseph J, Graves John A, Gruber Jonathan, and Kleiner Samuel A. 2015. "Measuring returns to hospital care: Evidence from ambulance referral patterns." *Journal of Political Economy*, 123(1): 170–214. [PubMed: 25750459]
- Doyle Joseph J, Ewer Steven M, and Wagner Todd H. 2010. "Returns to physician human capital: Evidence from patients randomized to physician teams." *Journal of health economics*, 29(6): 866–882. [PubMed: 20869783]
- Duggan Mark, Starc Amanda, and Vabson Boris. 2016. "Who benefits when the government pays more? Pass-through in the Medicare Advantage program." *Journal of Public Economics*, 141: 50–67.
- Einav Liran, Finkelstein Amy, and Mahoney Neale. 2022. "Producing Health: Measuring Value Added of Nursing Homes."
- Ericson Keith, and Starc Amanda. 2015. "Measuring Consumer Valuation of Limited Provider Networks." *American Economic Review Papers and Proceedings*, 105(5): 115–119.
- Finkelstein Amy, Gentzkow Matthew, and Williams Heidi L. 2019. "Place-based drivers of mortality: Evidence from migration." National Bureau of Economic Research.
- Finkelstein Amy, Gentzkow Matthew, and Williams Heidi. 2016. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics*, 131(4): 1681–1726. [PubMed: 28111482]
- Finkelstein Amy, Taubman Sarah, Wright Bill, Bernstein Mira, Gruber Jonathan, Newhouse Joseph P, Allen Heidi, Baicker Katherine, 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. [PubMed: 23293397]
- Garthwaite Craig, and Notowidigdo Matthew. 2019. "Plan Value-Added: Evaluating Medicaid Managed Care Plans Using Random Assignment." Work in Progress.
- Garthwaite Craig, Ody Christopher, and Starc Amanda. 2020. "Endogenous Quality Investments in the U.S. Hospital Market" National Bureau of Economic Research Working Paper 27440.
- Gaynor Martin, and Town Robert J. 2011. "Competition in health care markets." *Handbook of health economics*, 2: 499–637.
- Geruso Michael, and Layton Timothy. 2017. "Selection in Health Insurance Markets and Its Policy Remedies." *Journal of Economic Perspectives*, 31(4): 23–50. [PubMed: 29465215]
- 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. [PubMed: 29465214]
- Gruber Jonathan, and McKnight Robin. 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 Kolstad Jonathan. 2015. "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review*, 105(8): 2449–2500. [PubMed: 29546969]
- Handel Benjamin, Holmes Jonathan, Kolstad Jonathan, and Lavetti Kurt. 2019. "Insurer Innovation and Health Care Efficiency: Evidence from Utah." Work in Progress.
- Hull Peter. 2020. "Estimating Hospital Quality with Quasi-experimental Data."
- IBM. 2020. "IBM Micromedex RED BOOK." <https://www.ibm.com/products/micromedex-red-book>.
- Iowa State University, Iowa Community Indicators Program. 2022. "Urban Percentage of the Population for States, Historical." <https://www.icip.iastate.edu/tables/population/urban-pct-states>.

- 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. 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. 2019a. "10 Things to Know About Medicaid Managed Care." <https://www.kff.org/medicaid/issue-brief/10-things-to-know-about-medicaid-managed-care/>.
- Kaiser Family Foundation. 2019b. "Medicaid MCO Enrollment by Plan and Parent Firm." <https://www.kff.org/other/state-indicator/medicaid-enrollment-by-mco/?currentTimeframe=0&sortModel=%7B%22colId%22:%22State%22,%22sort%22:%22asc%22%7D>.
- Kwok Jennifer H. 2019. "How do primary care physicians influence healthcare." Evidence on practice styles and switching costs from medicare. URL: <https://static1.squarespace.com/static/5bd6632951f4d49caf7eed5/5/1564679228053>.
- Layton Timothy, Ndikumana Alice, and Shepard Mark. 2017. "Health Plan Payment in Medicaid Managed Care: A Hybrid Model of Regulated Competition." National Bureau of Economic Research Working Paper 23518.
- Layton Timothy, Maestas Nicole, Prinz Daniel, and Vabson Boris. 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, Newhouse Joseph P, Duan Naihua, Keeler Emmett B, and Leibowitz Arleen. 1987. "Health insurance and the demand for medical care: evidence from a randomized experiment." *The American economic review*, 251–277. [PubMed: 10284091]
- 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>.
- Medi-Cal Managed Care Division. 2012. "Statewide Collaborative Quality Improvement Project Reducing Avoidable Emergency Room Visits.", (June).
- National Committee for Quality Assurance (NCQA). 2022. "Adult and Child Health Care Quality." <https://www.medicare.gov/medicaid/quality-of-care/performance-measurement/adult-and-child-health-care-quality-measures/adult-core-set-reporting-resources/index.html>.
- New York State Department of Health. 2008–2012. "Claims and Eligibility Data." <https://health.data.ny.gov/>.
- Office of the New York State Comptroller. 2021. "Medicaid: Enrollment Growth, COVID-19 and the Future." <https://www.osc.state.ny.us/files/reports/pdf/medicaid-enrollment-growth-covid-19-and-the-future.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.
- Salop Steven, and Stiglitz Joseph E. 1982. "The theory of sales: A simple model of equilibrium price dispersion with identical agents." *The American Economic Review*, 72(5): 1121–1130.
- Schwartz Aaron L, Landon Bruce E, Elshaug Adam G, Chernew Michael E, and McWilliams J Michael. 2014. "Measuring low-value care in Medicare." *JAMA internal medicine*, 174(7): 1067–1076. [PubMed: 24819824]
- Song Zirui, Landrum Mary Beth, and Chernew Michael E. 2013. "Competitive bidding in Medicare Advantage: Effect of benchmark changes on plan bids." *Journal of health economics*, 32(6): 1301–1312. [PubMed: 24308881]

- Sparer Michael. 2012. "Medicaid Managed Care: Costs, Access, and Quality of Care." Robert Wood Johnson Foundation Research Synthesis Report 23.
- United Hospital Fund. 2004. "UHF Codes." <https://www1.nyc.gov/assets/doh/downloads/pdf/ah/zipcodetable.pdf>.
- United States Census Bureau. 2010. "Zip Code Tabulation Areas (ZCTAs)." <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html>.
- United States Census Bureau. 2019. "Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019." <https://www2.census.gov/programs-surveys/pepst/tables/2010-2019/state/totals/nst-est2019-01.xlsx>.
- 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 10.2139/ssrn.3544928.
- Wallace Jacob, McWilliams J Michael, Lollo Anthony, Eaton Janet, and Ndumele Chima D. 2022. "Residual confounding in health plan performance assessments: evidence from randomization in Medicaid." *Annals of internal medicine*.
- Washington Health Alliance. 2015. "Measure Specifications for potentially avoidable ER visits in 2015." <https://wahealthalliance.org/wp-content/uploads/2015/01/Measure-specifications-for-potentially-avoidable-ER-visits.pdf>.
- 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.

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



(b) Medicaid Spending as % of Total State Spending

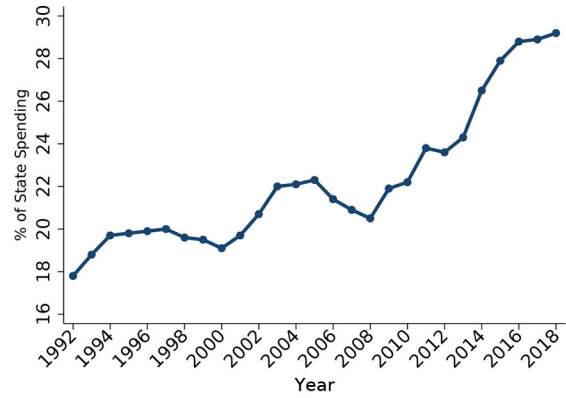


Figure 1: Trends in Medicaid Enrollment and Spending, 1992–2018

Note: Figure displays trends in Medicaid managed 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 Medicaid and Medicare Services and Mathematica Policy Research, 2016, 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.

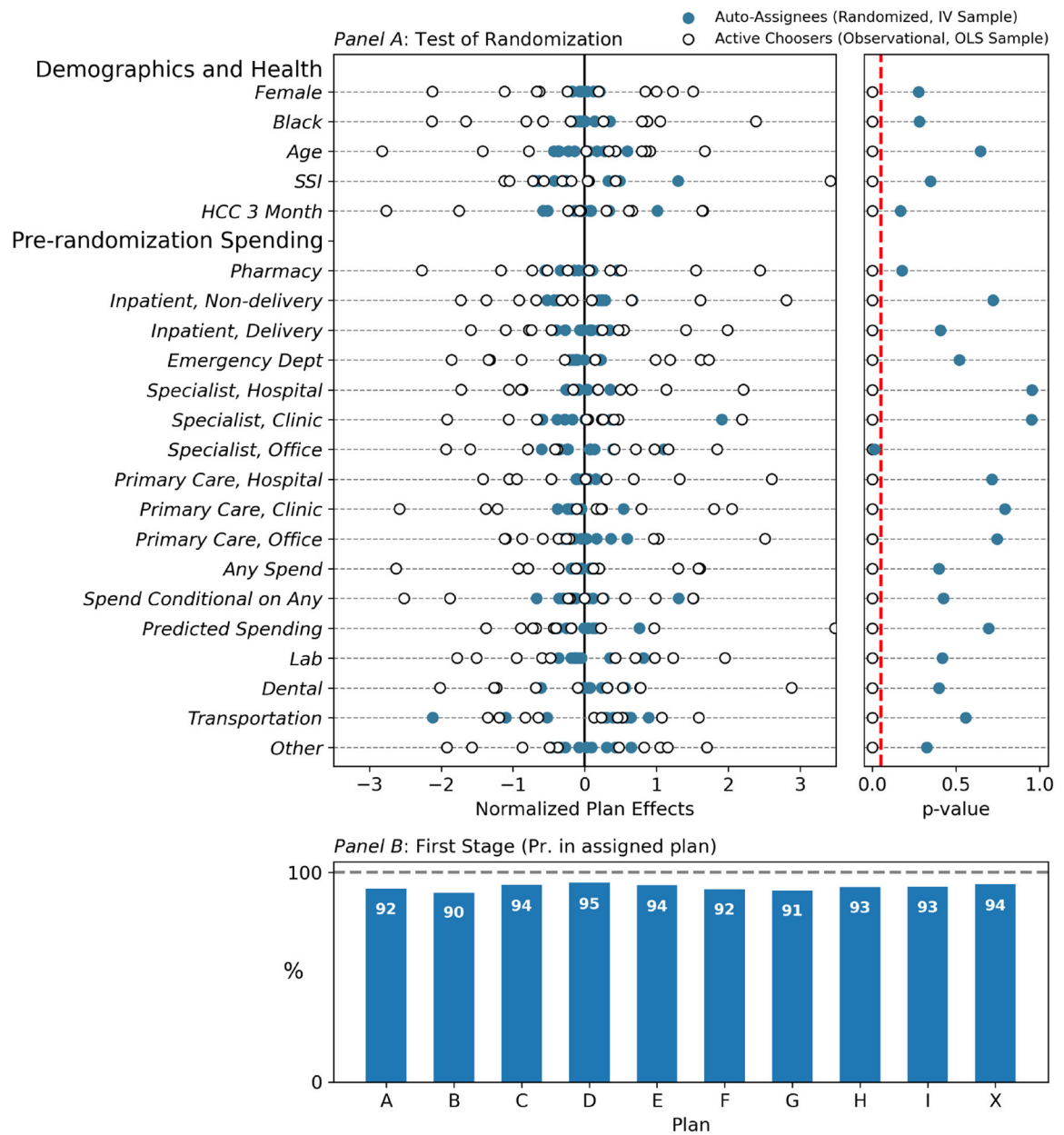


Figure 2: First Stage and Instrument Balance on Predetermined Characteristics
Note: Figure displays a balance test for the randomization in panel (a) and first stage regression coefficients in panel (b). Pre-determined characteristics include demographics and healthcare utilization in FFS Medicaid prior 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 A4. 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 A3 reports all first stage coefficients.

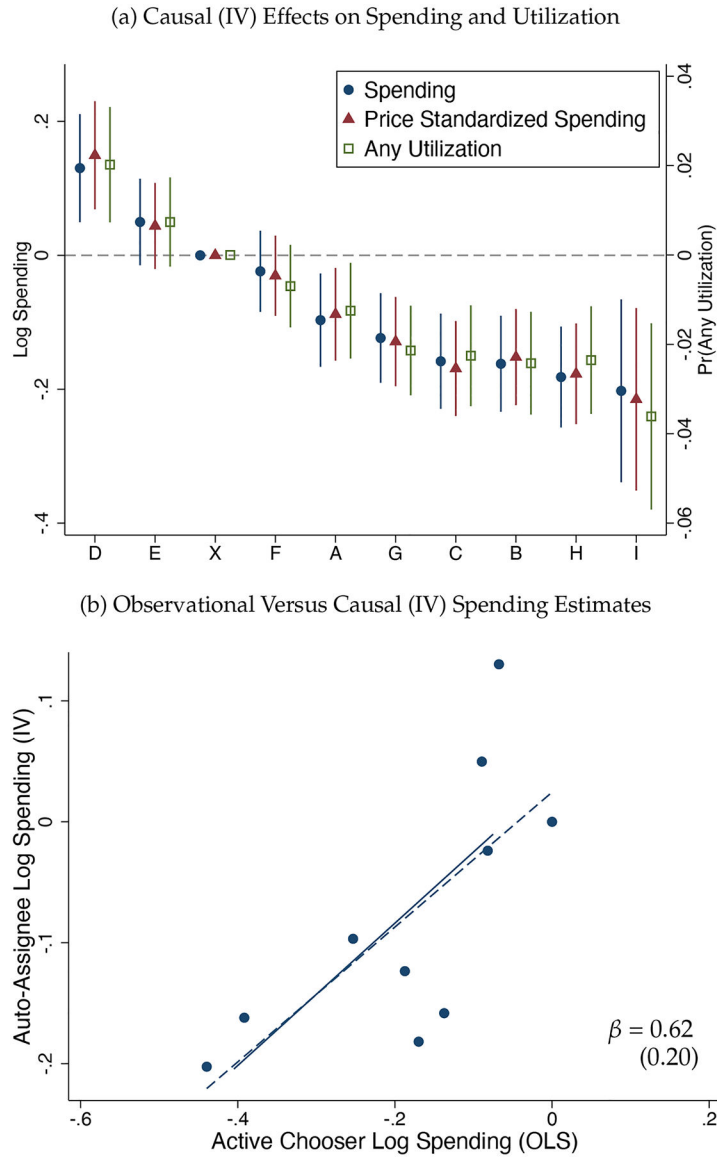


Figure 3:
 Main Results: IV Plan Effects on Healthcare Utilization
Note: Figure displays a main result of the paper—plan effects on healthcare utilization identified by random plan assignment. Panel (a) plots IV coefficients corresponding to Eq. 3, where the dependent variable is log(total healthcare spending + 1) and price-standardized spending 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 estimated in the active chooser sample. The dashed line is an OLS regression line fit to the ten points. The regression represented in the solid line uses an

empirical Bayes procedure to shrink coefficients prior to estimation. The slope of the solid line is reported. 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 A11 for tabular forms of these results and for complete details on the control variables and reweighting.

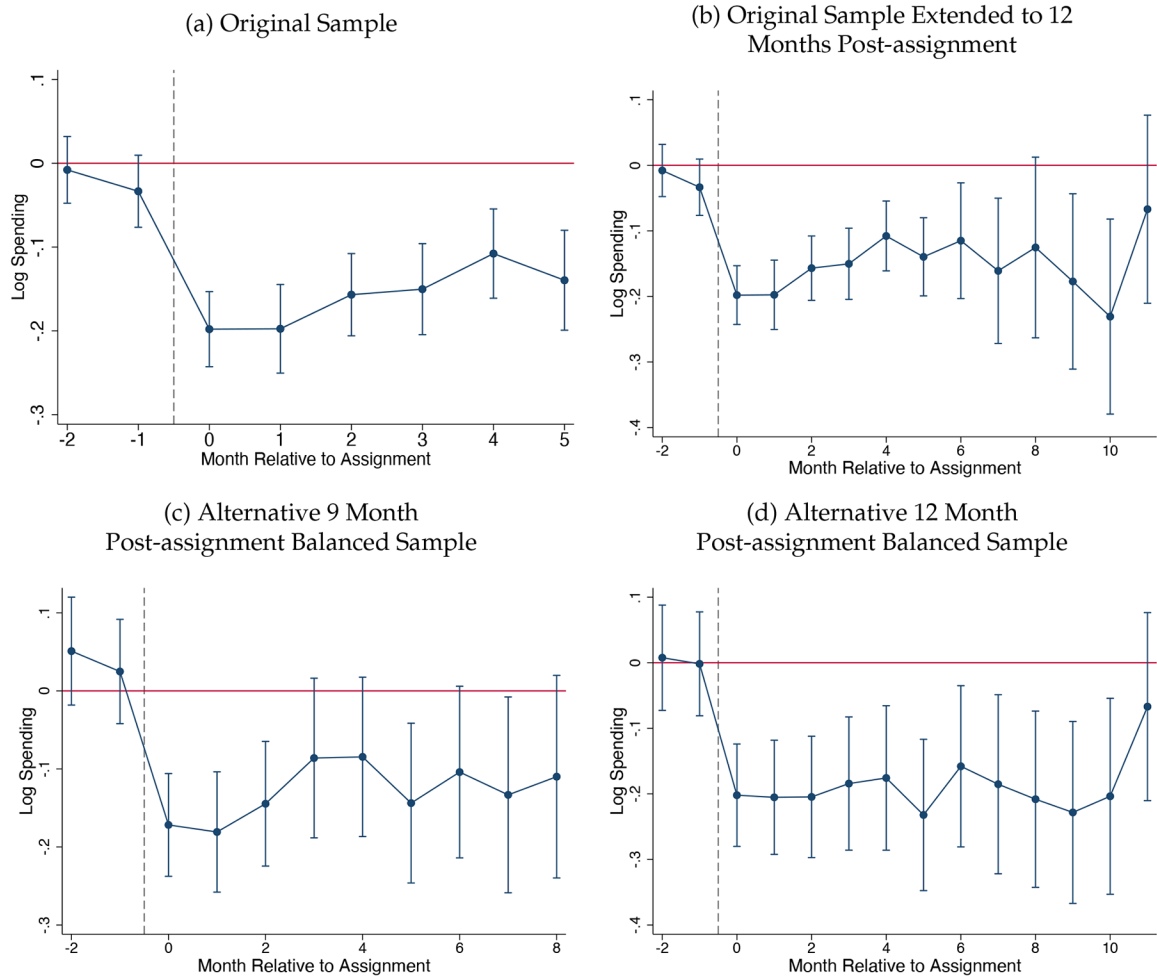


Figure 4: Persistence: Effects By Time Since Assignment to a Plan
Note: Figure displays results in the spirit of difference-in-difference event studies showing the spending impacts of being assigned to a low- versus medium-spending plan. As in Table A13, we divide the ten plans into three sets: low- (Plans A, B, C, G, H, I), medium- (Plans E, F, and X), and high- (Plan D) spending plans. Medium-spending plans are the omitted category and results for low-spending plans are shown. Event time (τ) is along the horizontal axis with month zero corresponding to the first month post-assignment. Using a modification of the IV regression in Equation (3), each point is estimated from a separate regression (one for each τ) of the form:

$$\log(\text{Spending} + 1)_{it}^{\tau} = \alpha^{\tau} + \phi_{e(i)}^{\tau} + \delta^{\tau} X_{it} + \lambda_{\text{low}}^{\tau} \widehat{\text{LowPlan}}_{it} + \lambda_{\text{high}}^{\tau} \widehat{\text{HighPlan}}_{it} + \epsilon_{it}^{\tau}.$$

We plot point estimates and 95% confidence intervals for $\lambda_{\text{low}}^{\tau}$. 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.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

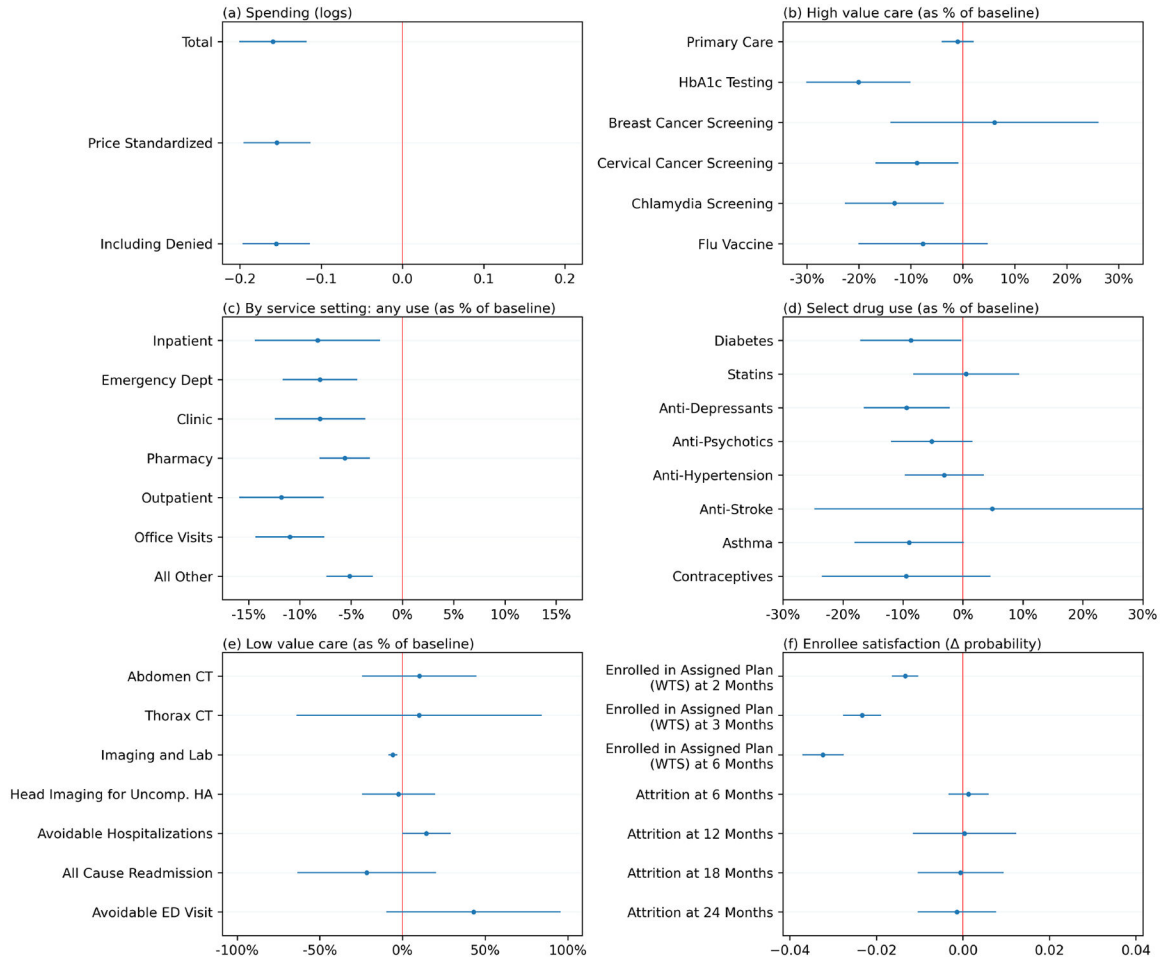


Figure 5: Low- Versus Medium-Spending Plan Effects Across Settings and Outcomes

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:

$$Y_{it} = \rho + \psi_{e(i)} + vX_{it} + \gamma_{Low} \widehat{Low\ Plan}_{it} + \gamma_{High} \widehat{High\ Plan}_{it} + \mu_{it}$$

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-stay; 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 A19, A20, A21, and A22.

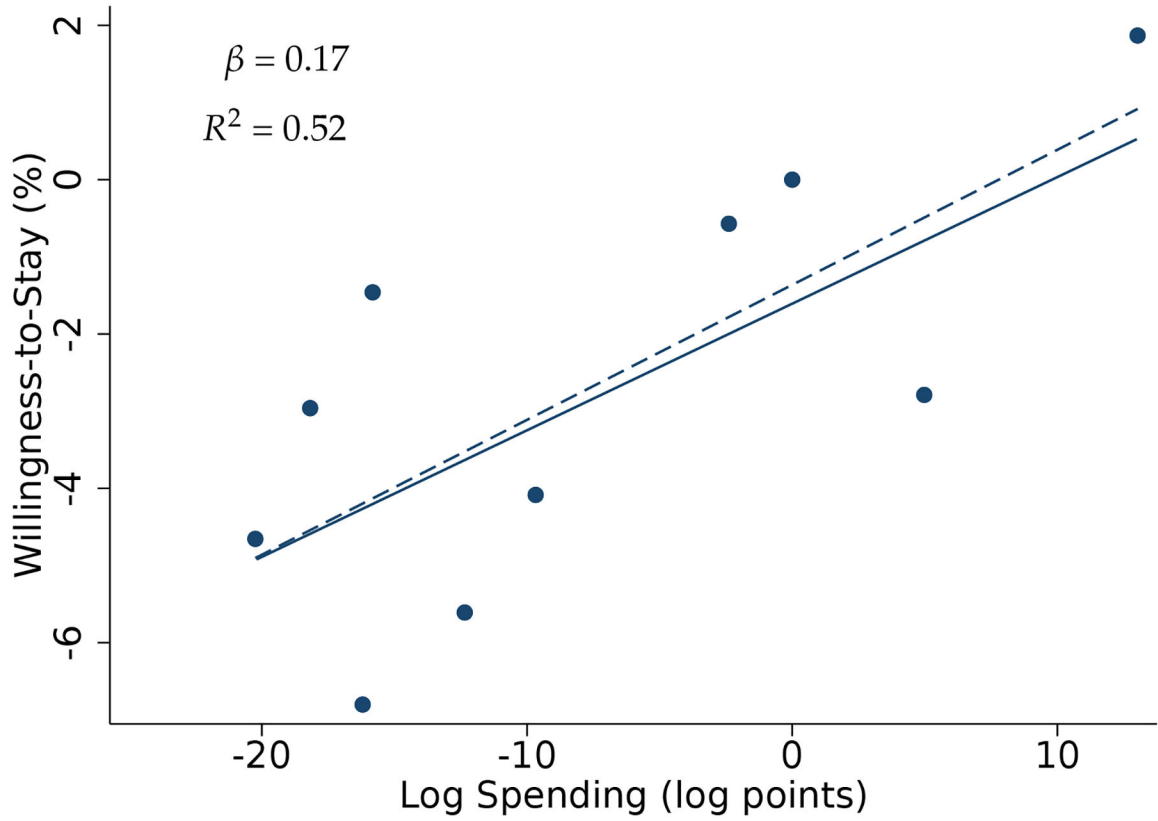


Figure 6:

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 dashed line is an OLS regression line fit to the ten points. The regression represented in the solid line uses an empirical Bayes procedure to shrink coefficients prior to estimation. The slope and R^2 from the shrunken estimation is reported.

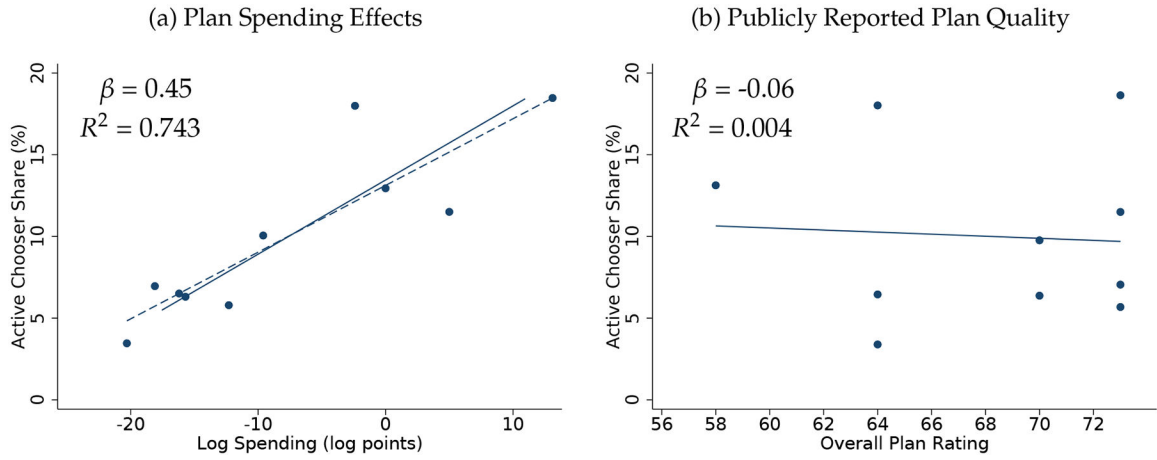


Figure 7:
 Plan Market Shares Versus Plan Spending Effects and Publicly Reported Plan Quality
Note: Figure shows the strong correspondence between active chooser share and plan spending (panel (a)) and the weak correspondence between active chooser share and publicly reported plan quality (panel (b)). 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 dashed line in panel (a) is an OLS regression line fit to the ten points. The regression represented in the solid line uses an empirical Bayes procedure to shrink the spending coefficients prior to estimation. The slope and R^2 from the shrunken estimation is reported. There is no shrinkage procedure used in panel (b).

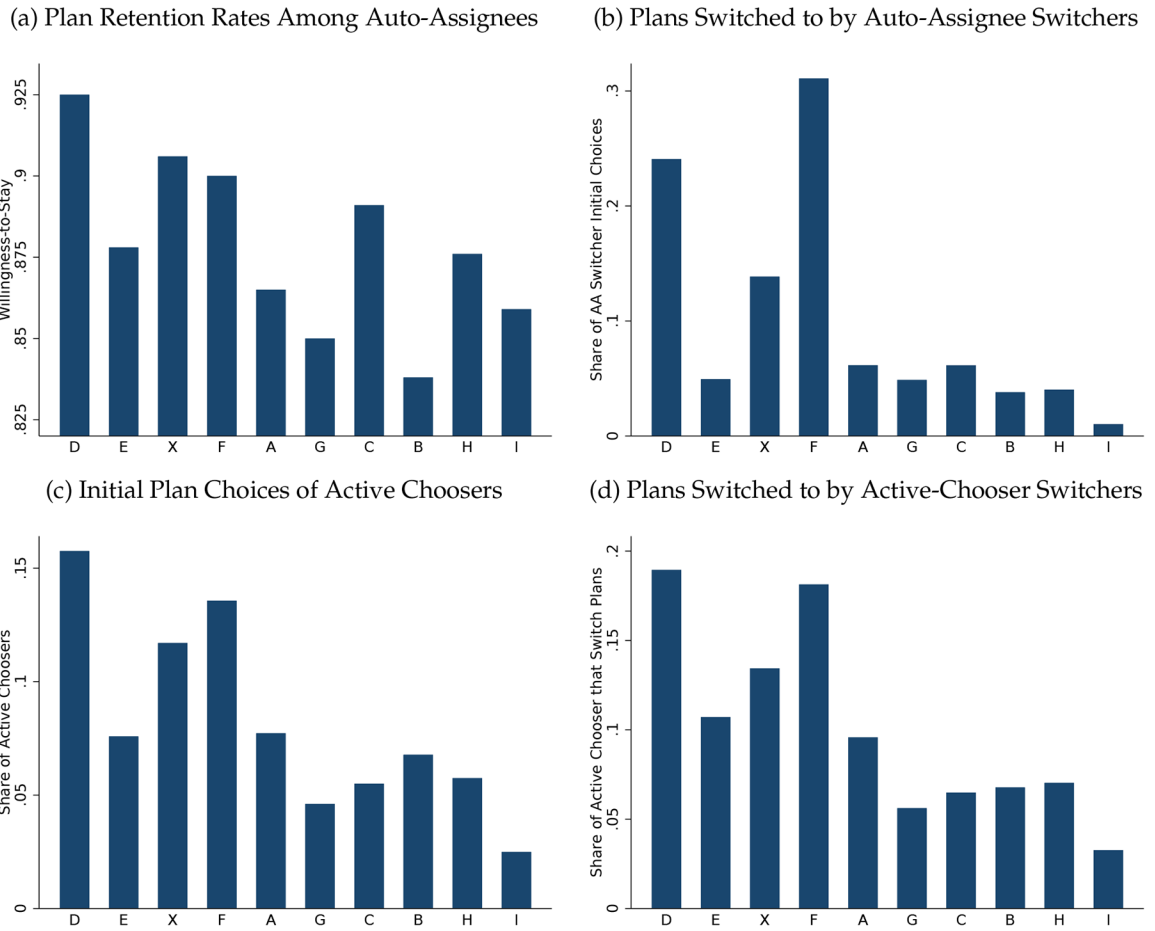


Figure 8:
The Highest Spending Plans Tend to Be Most Preferred

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)
<i>Demographics</i>			
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
<i>Healthcare Spending, \$ per enrollee-month</i>			
Total	510	2,877	393,570
Office Visits	21	165	393,570
Clinic	52	280	393,570
Inpatient	220	2,546	393,570
Outpatient	41	302	393,570
Emergency Dept.	16	100	393,570
Pharmacy	75	454	393,570
All Other	84	621	393,570
Enrollees with Any Spending (%)	34.87	47.65	393,570
Spending Conditional on Any (\$)	1,462	4,727	137,222
<i>Drug Days Supply, days per enrollee-month</i>			
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
<i>High-Value Care, per 1,000 enrollee-months</i>			
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
<i>Low-Value Care, per 1,000 enrollee-months</i>			
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
All Cause Readmission	0.29	18.59	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.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Main Results: Plan Effects on Spending and Plan Switching

Plan	Summary Statistics			Regression Results							
	Number of AutoAssignees (IV Sample)	% of Active Choosers Selecting Plan	For-profit	OLS Spending			IV Spending		Willingness-to-Stay		
				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?
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)		
D	2,621	18.47		0.165** (0.027)	-0.111** (0.023)	-0.064 (0.055)	0.171** (0.050)	0.130** (0.041)	0.020** (0.007)	0.014** (0.004)	0.019** (0.005)
E	6,763	11.51		0.076* (0.033)	-0.084** (0.027)	-0.065 (0.055)	0.058 (0.040)	0.050 (0.033)	0.007 (0.005)	-0.025** (0.003)	-0.028** (0.004)
X	8,698	12.94									
F	8,057	17.99		0.286** (0.030)	-0.126** (0.025)	-0.040 (0.048)	-0.011 (0.036)	-0.024 (0.031)	-0.007 (0.005)	-0.001 (0.003)	-0.006 (0.004)
A	8,512	10.06		-0.265** (0.035)	-0.290** (0.029)	-0.276** (0.057)	-0.101* (0.042)	-0.097** (0.035)	-0.012* (0.005)	-0.029** (0.004)	-0.041** (0.004)
G	8,444	5.79		-0.003 (0.038)	-0.225** (0.031)	-0.214** (0.068)	-0.134** (0.041)	-0.123** (0.034)	-0.021** (0.005)	-0.041** (0.004)	-0.056** (0.004)
C	6,198	6.31		0.066+ (0.037)	-0.142** (0.031)	-0.188** (0.067)	-0.166** (0.044)	-0.158** (0.036)	-0.023** (0.006)	-0.010* (0.004)	-0.015** (0.005)
B	7,815	6.51	✓	-0.551** (0.046)	-0.459** (0.035)	-0.394** (0.067)	-0.178** (0.042)	-0.162** (0.036)	-0.024** (0.006)	-0.047** (0.004)	-0.068** (0.004)
H	7,066	6.96	✓	-0.100* (0.042)	-0.125** (0.034)	-0.152* (0.070)	-0.158** (0.046)	-0.182** (0.038)	-0.024** (0.006)	-0.020** (0.004)	-0.030** (0.005)
I	1,417	3.46	✓	-0.522** (0.046)	-0.375** (0.040)	-0.423** (0.073)	-0.165+ (0.084)	-0.202** (0.069)	-0.036** (0.011)	-0.030** (0.006)	-0.047** (0.008)
Mean (spend displayed in dollars)				466.202	466.202	462.263	509.740	509.740	0.349	0.930	0.906
County × Year × Month FEs				X	X	X	X	X	X	X	X
Person-Level Controls					X	X		X	X	X	X
First Stage F-Test P-Value							<0.001	<0.001	<0.001		
F-Test P-Value				<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Plan Effect SD				0.265	0.135	0.140	0.112	0.105	0.016	0.018	0.026
Corrected SD				0.263	0.132	0.126	0.102	0.098	0.015	0.018	0.025
Obs: Enrollees										65,595	65,595
Obs: Enrollee X Months				592,692	592,692	392,026	393,570	393,570	393,570		

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 A3 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 × months in columns 3 through 8 and enrollees in columns 9 and 10. OLS regressions include only active-choosers; see Table A24 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 × year × 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.

⁺ $p < 0.1$,

^{*} $p < 0.05$,

^{**} $p < 0.01$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript