The Welfare Effects of Accountable Care Organizations

James Gluzman *

September 30, 2025

Abstract

This paper studies the welfare effects of Accountable Care Organizations (ACOs), a policy whose goal is to reduce healthcare spending for Traditional Medicare patients. ACOs are groups of healthcare providers whose performance is evaluated by comparing a spending benchmark to the total outpatient and inpatient spending of assigned patients. Spending below the benchmark results in a financial reward given to the ACO, while spending above the benchmark results in a penalty. The welfare impact of this policy is theoretically ambiguous, as lower spending may reduce quality of care. To quantify these effects, I develop a structural model of supply and demand in the Medicare outpatient facility market where facilities compete in outpatient spending and quality. The model incorporates endogenous quality provision and spillovers onto Medicare Advantage patients, whose healthcare spending is not directly targeted by the program. I estimate the model using the universe of hospital-based claims from New York State. A counterfactual simulation that removes ACOs from the market shows that ACOs increase welfare by $1.09 billion. Spillovers onto Medicare Advantage patients account for 32% of the consumer welfare gains.

Keywords: Medicare, Healthcare Policy, Healthcare spending, Quality, Cost Containment.

JEL Classification Codes: H2, H3, I11, I13, I18, L1, L2

*Gluzman: Economics Department, Stony Brook University, james.gluzman@stonybrook.edu. I am immensely grateful to Yiyi Zhou, Steven Stern, Ting Liu and Kathleen McGarry for their guidance and support. This work has benefited from discussions with Hugo Benitez-Silva, Marc Claveria Mayol, Gautam Gowirsanakaran, Ashley Langer, Robert Millard, Mark Montgomery, Takahiro Moriya, David Rivers, Eran Shmaya, Ashley Swanson and Yi Xin. All errors are my own.

1 Introduction

Healthcare spending in the Medicare program is both substantial and growing.[1] In 2023, annual spending per beneficiary averaged roughly $13,000, adding up to nearly $1 trillion each year, or about 3% of U.S. GDP (CMS, 2025c). Policymakers have expressed concern over the sustainability of this high level of spending and have called for reforms to reduce costs. However, lowering healthcare spending is not without risks: reductions in spending may come at the expense of the quality of care patients receive (Cutler, 1995; Cooper et al., 2019). As the U.S. population continues to age and medical costs rise, the challenge is not only to contain spending on Medicare but to identify mechanisms that improve the burden of high healthcare spending while safeguarding or even enhancing quality. Understanding the effects of policies aimed at reducing healthcare spending requires considering their potential implications for patient care quality.

This paper focuses on Medicare Accountable Care Organizations, hereby referred to as ACOs: a supply-side mechanism designed to incentivize reductions in healthcare spending while maintaining or improving quality of care. Under the program, groups of healthcare providers, such as physicians and hospitals, voluntarily enter into a contract with Medicare, agreeing to be held accountable for both the healthcare spending and the quality outcomes of an assigned group of patients. Each ACO's performance is assessed by comparing the realized outpatient and inpatient spending of its assigned patients to a pre-determined benchmark. Maintaining a level of total spending below the benchmark generates a financial reward for the ACO, whereas spending above the benchmark results in a financial penalty. In both cases, the resulting payment or penalty is distributed among the participating providers of the ACO. Healthcare providers face two key incentives under this program. First, they are encouraged to reduce outpatient spending for their assigned patients by lowering the quantity of outpatient services administered. Second, they are incentivized to reduce inpatient spending by improving quality of care, which lowers the likelihood that patients require inpatient treatment.

The welfare implications of ACOs are theoretically ambiguous. While the program is designed to incentivize providers to reduce healthcare spending through reductions in quantity of services administered to patients and improvements in quality of care, these incentives may generate unintended consequences. On one hand, efforts to reduce outpatient spending can translate into lower revenue for healthcare providers that can cause disinvestment in quality of care, even under incentives to improve quality (Cutler, 1995; Doyle et al.,

[1] Medicare is a federal health insurance program that primarily covers U.S. citizens and permanent residents aged 65 and older as well as younger individuals with certain disabilities or end-stage renal disease.

2015; Cooper et al., 2019). On the other hand, incentives to improve quality may drive up healthcare spending to off-set quality investment costs, even with incentives to reduce healthcare spending, creating a case of quality overprovision (Crawford et al., 2019). In addition, the overall effects of ACOs extend beyond participating providers. Competition with ACO participants can shift the quantity of services and quality of care chosen by non-participating providers. Moreover, patients excluded from ACO performance metrics can experience changes in the care they receive. The effects of ACO incentives on participants' outpatient spending and quality, as well as on competition and spillovers, make the policy's welfare consequences uncertain.

To quantify the welfare effects of ACOs, I construct a structural model of supply and demand in the Medicare outpatient facility market, following Berry et al. (1995), Fan (2013) and Crawford et al. (2019). On the supply side, multi-facility hospital systems simultaneously compete in the dimensions of quantity of outpatient services and quality of care, where quantity of services maps onto outpatient spending and costs are increasing in quality. Systems in an ACO incorporate the ACO's performance into their objective function. On the demand side, patients who are heterogeneous in location, diagnoses and insurance coverage choose the outpatient facility that maximizes their utility, observing the quality and spending decisions of outpatient facilities.

The supply side models multi-facility outpatient hospital systems as spatially differentiated products competing in both quantity of outpatient services and quality of care. It incorporates the incentives of ACOs, where participating systems evaluate the ACO's performance by comparing the total spending of their assigned Traditional Medicare patients against the ACO's benchmark. Participation in an ACO creates incentives for systems to reduce outpatient spending through lowering the quantity of outpatient services administered to patients and to decrease inpatient spending by increasing quality of care where higher quality of care reduces inpatient admission rates. By participating in an ACO, systems have an increased marginal disutility from increasing the quantity of services administered and have an increased marginal benefit from increasing quality of care, compared to non-participating systems. Importantly, systems also account for the presence of Medicare Advantage patients in their objective function when making decisions. Although these patients are not directly targeted by ACO incentives, the changes in quantity of services and quality of care induced by ACO participation generate spillover effects onto Medicare Advantage patients, affecting their care and outcomes. Systems that do not participate in an ACO respond strategically to the choices of ACO participants through competition.

The demand model incorporates three important features of the Medicare outpatient hospital market. First, I allow for price sensitivity to vary across insurance categories, as differences in coverage generosity affect price responsiveness (Ho and Pakes, 2014). Second, I incorporate patient preferences over travel distance, where consumers may incur a disutility for traveling from their residence to outpatient facility of choice, which allows for spatial competition among different facilities (Kessler and McClellan, 2000; Gowrisankaran et al., 2015; Ho and Lee, 2017). Third, I account for patient inertia in returning to past-visited facilities, and steering of patients between facilities under the same hospital system or ACO (Ho and Pakes, 2014; Prager, 2020; Chernew et al., 2021). Incorporating inertia and steering in the demand model introduces a potential trade-off between facility market power and potential efficiency in steering patients to facilities with lower spending and higher quality.

The empirical setting to which I apply the model is the New York State outpatient facility market. I estimate the parameters of the demand model using the Statewide Planning and Research Cooperative System (SPARCS) data set, which contains the universe of administrative outpatient and inpatient hospital-based claim-level data on patient choices of outpatient facilities in New York State. The claims data contain patient choices of outpatient facilities along with individual-level characteristics such as residential zip code, diagnoses and insurance coverage. Using system profit-maximizing optimality conditions and the estimated demand system, I recover each facilities' quality- and non-quality-related marginal costs.

The estimates of the demand model are reasonable. I find that patients prefer higher quality outpatient facilities, where patients' willingness to pay for a one standard deviation increase in quality is between \$45 and \$48 (Tay, 2003; Gaynor et al., 2016; Hackmann, 2019). In addition, Medicare patients dislike higher spending facilities, which is consistent with the results of Baker et al. (2016). Moreover, patients with higher generosity insurance coverage are less sensitive to higher spending facilities, compared to those with less generous coverage (Gowrisankaran et al., 2015; Hoffman, 2015; Hackmann, 2019). Consistent with existing literature on patient choice of healthcare providers, patients dislike facilities that are located a greater distance from their residential zip code (Kessler and McClellan, 2000; Gowrisankaran et al., 2015). The model estimates imply that patients' willingness to pay for reducing travel distance by one mile ranges between \$58 and \$62. There is a significant inertia effect, where patients are approximately 4.5 times more likely to return to facilities they previously visited (Prager, 2020; Shepard, 2022). Additionally, steering effects are strong within systems and ACOs (Baker et al., 2016; Chernew et al., 2021). Patients who have previously visited a facility in a given system are about 4.7 times more likely to choose another facility within the same system for future care. Similarly, patients are about 4.8 times more likely to seek care at another facility within the same ACO.

To evaluate and quantify the welfare effects of ACOs, I use the model to simulate a counterfactual that removes ACO incentives from systems' objective function. Taking into account changes in consumer welfare, producer welfare and government spending, ACO incentives increase welfare by $1.09 billion. The counterfactual simulation shows that ACOs incentivize facilities to decrease outpatient spending by 4.7% per patient and increase quality by 2.1%. The decrease in outpatient quantity of services lowers annual outpatient spending by $546 per patient, while the increase in quality of care decreases annual inpatient spending by $260 per patient.

There are two mechanisms of note that drive this result. First, the incentives of ACOs incentivize participating systems to reduce the outpatient spending of its facilities while increasing quality of care. The incentive to improve quality introduced by ACOs outweighs the reduction in the marginal benefit caused by lower outpatient spending, resulting in equilibrium outcomes with both lower outpatient spending and higher quality of care. Second, non-participating systems respond competitively by reducing outpatient spending and increasing the quality of care at their own facilities to compete with ACO participants. Consumer welfare, measured as the combined welfare of all patients in both Traditional Medicare and Medicare Advantage, rises by $0.25 billion. Decomposing the gains by insurance type, Traditional Medicare beneficiaries account for $0.17 billion in welfare gains, and Medicare Advantage consumers benefit $0.08 billion through spillover effects. Producer welfare increases by $0.11 billion overall, where ACO participants gain $0.15 billion and non-participants experience a loss of $0.04 billion. Government spending decreases by $0.73 billion due to the lower healthcare spending of Traditional Medicare and Medicare Advantage patients.

Literature This paper contributes to the literature on supply-side incentives and organizational structures in healthcare utilization markets. Ellis and McGuire (1986) develops a model showing that prospective payment systems can lead physicians to under-provide services. Empirically, Cutler (1995) finds that hospitals receiving lower payments under the Medicare prospective payment system delivered lower quality care as measured by higher mortality and readmission rates. In an incentive structure similar to ACOs, Gaynor et al. (2004) demonstrates that incentives to reduce medical spending in Health Maintenance Organizations lead providers to cut spending by approximately 5%. Building on the work of provider responses to spending-reduction incentives, Ho and Pakes (2014) shows that providers' heightened sensitivity to patient spending, via participation in ACOs, affects their referral choices. Aswani et al. (2019) calculates optimal contracts for ACO participants to improve cost savings. This paper is most closely related to Reddig (2024), who quantifies the spending-quality trade-off for ACOs. Unlike Reddig and other prior studies, this paper models the decisions of ACO participants and non-ACO participants simultaneously, which captures both the competitive spillover effects on non-participating providers and the impacts on patients who are not directly considered in ACO performance evaluations but are nonetheless affected by these incentives.

Second, this paper contributes to a large literature on price and quality competition in the healthcare market. Most prior research, such as Gaynor and Vogt (2003), models oligopoly competition solely over price in the inpatient hospital market. Similarly, Gowrisankaran et al. (2015) and Ho and Lee (2017) model hospital bargaining over prices with insurers in the inpatient hospital market. Other empirical studies, such as Tay (2003) and Gaynor et al. (2016), examine the role of differentiation over quality in healthcare markets. Rarely is the relationship between medical spending and quality discussed in an oligopoly setting. Hackmann (2019) develops a model of price and quality competition in the nursing home industry to quantify the welfare effects of higher Medicaid reimbursement rates designed to incentivize quality improvements. Similarly, Doyle et al. (2015) and Cooper et al. (2019) show the casual relationship between higher hospital prices and higher quality of care. This paper contributes to the literature by using a structural model to simultaneously model spending and quality decisions in the outpatient healthcare utilization market, considering the full set of competing systems and examining the equilibrium under alternative payment incentives rather than focusing solely on decision-making within individual systems.

Lastly, this paper contributes to the industrial organization literature on endogenous product characteristics in oligopoly markets. Fan (2013) builds on the standard oligopoly price competition framework of Berry et al. (1995) by endogenizing quality in the newspaper market and demonstrating the consequences of ignoring endogenous product attributes in oligopoly settings. Similarly, Crawford et al. (2019) shows that market structure and competitive forces can lead to overprovision of quality at the expense of higher prices. Gluzman and Zhou (2025) illustrates that regulations in product design in oligopoly markets with selection can have ambiguous welfare effects. Building on this literature, this paper shows that the interaction of incentive structures and competition can improve welfare, but overly strong incentives to increase quality may generate welfare losses through overprovision.

Organization The rest of the paper is organized as follows. In Section 2, I describe the institutional setting and present the data. Section 3 characterizes the empirical model, followed by a description of the estimation methods in Section 4. Section 5 reports the estimation results. Section 6 conducts counterfactual analyses to quantify the welfare impact of the ACO program. Section 7 provides concluding remarks. Additional details on data construction and industry background, along with additional Figures and Tables, are provided in the Appendix.

2 Institutions and Data

In this section, I provide institutional background on Medicare and ACOs. Then, I discuss the context relevant to my empirical application. Finally, I describe the data used for estimation and present summary statistics.

2.1 Medicare

Medicare is a federal health insurance program that primarily covers individuals who are age 65 and older in addition to those with disabilities or end-stage renal disease. In 2023, Medicare covered approximately 65 million beneficiaries with an average per beneficiary spending of about \$13,000. This translates to an aggregate annual spending of \$1 trillion, roughly 3% of the U.S. GDP (CMS, 2025c)

Medicare spending has grown steadily over the past several decades, with average annual growth rates of around 5%, driven by rising healthcare costs and a growing beneficiary population. This persistent growth has led to repeated calls from government officials and policymakers to constrain spending and improve program efficiency. In response, a variety of policy measures have been attempted, such as the Medicare Advantage program and the prospective payment system. While some of these policies have produced modest savings and improvements in care coordination, aggregate Medicare spending continues to increase. Under Medicare Advantage, costs per beneficiary have generally been slightly lower than under traditional fee-for-service Medicare, largely due to plan incentives to manage utilization, coordinate care, and negotiate provider payments. However, this cost advantage may also be partly due to Medicare Advantage plans engaging in "cream skimming" where insurers select healthier beneficiaries who require less care (Gluzman and Zhou, 2025). Prospective payment initiatives, which set a single payment for all services related to a hospitalization or clinical episode, have also reduced costs for certain procedures by encouraging providers to eliminate unnecessary services, improve care transitions, and manage post-acute care more efficiently. Evidence suggests that quality under these prospective payments has been maintained or modestly improved, particularly for readmission rates and complication rates (Cutler, 1995; Doyle et al., 2015).

Medicare reimburses healthcare providers through two different reimbursement models, contingent on whether the patient's care is outpatient or inpatient. Outpatient providers are reimbursed through the fee-for-service (FFS) model, where providers receive payments for each specific service they render. The government establishes set payment rates for each service, but providers have the flexibility to decide the types of services to offer and their quantity. Conversely, for inpatient hospital care, Medicare uses the Diagnosis Related

Group (DRG) system, which reimburses hospitals based solely on the patient's diagnosis. As a result, hospitals receive a predetermined payment amount, independent of the quantity, type of services or quality of care administered during the patient's stay.

2.2 Outpatient Facilities

There are several important reasons for my focus on the Medicare outpatient service market. To begin with, provider choices significantly shape outpatient spending due to the fee-for-service structure, where providers decide on the type and quantity of services provided. Additionally, because Medicare beneficiaries incur out-of-pocket expenses for outpatient services, this setting is important for managing the cost burden of Medicare beneficiaries.[2] Lastly, the quality of care received at the outpatient level can notably reduce the chances of inpatient hospitalization, which indirectly affects inpatient spending and health outcomes by altering the extensive margin of transition from outpatient care to inpatient care.

2.3 Accountable Care Organizations

Given the substantial costs associated with Medicare, both per beneficiary and in aggregate, policymakers have introduced Accountable Care Organizations (ACOs) as a supply-side incentive designed to reduce Medicare spending while maintaining high quality of care.[3] Participation in the ACO program by healthcare providers is completely voluntary. ACOs operate by holding a group of healthcare providers (e.g. physicians, physician groups, hospitals) accountable for the total spending for a specified group of patients. Participants in an ACO are incentivized by the performance of the ACO where performance is calculated based on the difference between a predetermined benchmark and the actual spending of the patients assigned to the ACO.

The performance of an ACO depends on the group of patients assigned to the ACO. A beneficiary is assigned to an ACO if that ACO accounts for the highest share of their primary care spending over all other ACOs. If a patient receives the plurality of primary care from providers that do not participate in an ACO, then the patient is not assigned to any ACO.[4] This assignment determines which patients are included in the ACO's spending

[2]Individuals enrolled in Traditional Medicare who do not enroll in supplementary insurance (e.g. Medigap or employer-sponsored supplementary insurance) face a 20% coinsurance rate for all outpatient services. On the other hand, inpatient services are fully covered for the first 60 days of a hospital stay. This typically results in very little out-of-pocket costs for inpatient services.

[3]I focus on Medicare ACOs under the Medicare Shared Savings Program. Similar incentives exist in the commercial insurance industry, which are known as commerical ACOs.

[4]See Appendix A.1 for more details on assignment rule methodology.

calculation, as well as the total benchmark the ACO receives.[5] The benchmark is determined by multiplying the number of assigned patients by a per-patient benchmark where the per-patient benchmark is the average annual Medicare spending per beneficiary in the provider's counties of service from the previous year. The ACO's spending aggregates the spending of patients assigned to the ACO, which includes spending accumulated by providers that do not participate in the ACO. This causes ACO participants to internalize spending decisions made by non-ACO providers. Moreover, outpatient providers in an ACO are responsible for the inpatient expenses of patients assigned to the ACO, a feature not present in the conventional Fee-For-Service system.

The ACO's performance is the difference between its benchmark and realized spending of patients assigned to the ACO, including spending accumulated by these patients under providers that do not participate in the ACO. The government retains a portion of any savings achieved, while the remainder is distributed among the participants of the ACO. ACOs with a single participant retain the entirety of the performance after the government keeps its share, while ACOs with multiple participants divide the performance according to a sharing rule negotiated prior to the beginning of the performance period. Participants of the ACO receive the shared performance of the ACO in addition to conventional Fee-For-Service revenue.

Consider a simple example of ACO performance with two providers: provider A and provider B with one patient in the market who is assumed to be assigned to the ACO. Assume that the ACO retains the full share of performance and does not split the performance with the government. The performance of the ACO is

$$\pi = \text{Benchmark} - \text{Outpatient Spending} - \text{Inpatient Spending}. \tag{1}$$

The performance of the ACO is captured by the difference between the ACO's benchmark and ACO's total spending (outpatient spending and inpatient spending). Then, the two providers split the performance according to a pre-negotiated sharing rule. Provider A receives $w \times \pi$, while provider B receives $(1 - w) \times \pi$, where $w \in (0, 1)$ is the sharing rule between the two participating providers (A and B). If the performance is negative, then both participating providers receive a penalty rather than a bonus.

In the example ACO performance shows in equation 1, outpatient spending is increasing

[5] Patient assignment is not an active choice by the patient; that is, the patient does not decide which ACO the be assigned to directly. Assignment occurs at the end of each performance year and is used strictly for determining the ACO's performance. Furthermore, patients do not face any restrictions on provider choice if they are assigned to a particular ACO. In other words, a patient may visit providers outside of the ACO to which they are assigned.

in quantity of outpatient services administered to patients and inpatient spending is decreasing in quality of care. Both providers in the ACO are incentivized to reduce quantity of services in order to reduce outpatient spending. In addition, both providers in the ACO are incentivized to increase quality of care in order to reduce inpatient spending.

The performance of the ACO introduces multiple noteworthy incentives. First, a participating outpatient provider who generates revenue from outpatient spending now instead receives a penalty for the outpatient spending of patients assigned to the ACO. Second, a participating outpatient provider now receives penalties for inpatient spending of patients assigned to the ACO, which is a financial incentive not present in the standard reimbursement system. Third, the ACO receives a benchmark payment for each patient assigned to the ACO. One can think of this benchmark as a “subsidy” given to the ACO for taking on accountability, for each assigned patient. Fourth, participating providers do not know the identity of patients assigned to the ACO upon making spending and quality choices but have an expectation over patient assignment.[6] Finally, participating outpatient providers internalize the spending and quality decisions of all providers in the market, including providers that do not participate in the ACO. Because the ACO performance depends on total outpatient and inpatient spending for an assigned patient across all providers the patient visited (regardless of whether or not these providers participate in the ACO), the participating outpatient provider now directly internalizes the choices of competing providers.

As of 2019, there are 487 Medicare ACOs across the United States, with approximately 10.4 million beneficiaries (16% of Medicare Beneficiaries) being assigned to an ACO. Regarding hospital system participation, approximately 30% of U.S. participated in an ACO by 2019. The ACO program has experienced significant growth since its start, expanding from 114 ACOs in 2012 to 487 in 2019.

2.4 Empirical Setting & Data

My empirical setting is the hospital outpatient market for Medicare beneficiaries in New York State during the period of 2017 to 2019. During this time frame, there were 226 hospital-based outpatient facilities with the involvement of 17 distinct ACOs. The empirical analysis utilizes administrative claims data, which I merge with facility-specific information from various sources to build a detailed dataset on patient usage, facility characteristics and the market structure.

[6] This feature of uncertainty over patient assignment is not captured in the simplified example of ACO performance above, but is included in the empirical model.

Individual Level Claims Data

I utilize administrative records from the Statewide Planning and Research Cooperative System (SPARCS) encompassing all hospital-based outpatient and inpatient claims in New York State. This dataset provides comprehensive details at the claim level for hospital-based providers and facilities throughout the state.

At the individual level, I observe each patient's age, gender, and residential zip code. Information on the patients' zip code allows me to factor in patient preferences over travel distances between the patients' residence and facilities in the demand model. At the outpatient claim level, I observe patients' choice of facility, diagnoses, insurance details, admission and discharge dates alongside the payment received per visit. I assign each diagnosis to one of 18 mutually exclusive categories using the Clinical Classifications Software (CCS).[7] Insurance information is used to account for variations in price sensitivity between different types of insurance. Admission and discharge dates are used to determine the prior use of facilities for each patient, which is integrated into preferences captured in the demand model. For evaluating quality of care, I link outpatient discharge dates with subsequent inpatient admissions for the same patient occurring within 30 days, identifying transitions from outpatient to inpatient care. Lastly, I construct a measure of average quantity of services at the facility-year level by combining information on payments and diagnoses with the appropriate reimbursement rates for each visit.

My micro-level sample focuses on claims covered by Medicare, including both Traditional Medicare and Medicare Advantage enrollees. The final sample consists of 2,533,354 person-years and includes 16,212,409 outpatient claims. Table 1 provides summary statistics of the individuals and claims from the claims data.

The average age in the sample aligns with the Medicare demographic at around 70 years, with females constituting 58% of the sample. Examining insurance coverage, 64% of the individuals are covered by Traditional Medicare, while the remaining 36% are enrolled in a Medicare Advantage plan. Moreover, 25% of individuals are also on Medicaid, which can be combined with any Medicare plan type. Examining the claim level data, patients travel an average of approximately eight miles to visit their chosen healthcare provider, indicating a preference against long travel distance. The average payment for an outpatient service is approximately $1,900. Examining health outcomes, 4% of outpatient services follow up with an inpatient admission within 30 days of discharge. Lastly, 45% of outpatient visits are conducted by a provider that participates in an ACO.

[7]See Appendix A.2 for a description of the CCS diagnosis categories and their frequencies.

Table 1: Summary Statistics for Individual and Claim Level Data

Panel A: Individual-Level	Mean	Std. Dev.
Age	70.48	(12.90)
Female	0.58	
Traditional Medicare	0.64	
Medicare Advantage	0.36	
Medicaid	0.25	
Annual Outpatient Visits	7.09	(13.49)
Annual Inpatient Visits	1.21	(0.52)
Panel B: Claim-Level		
Distance Traveled	7.93	(12.29)
Outpatient Payment	1,901	(3,148)
Inpatient within 30 days	0.04	
Treated by ACO Participant	0.45	

Notes: Panel A reports statistics at the individual-year level (N = 2,533,354 patient-years). Panel B reports statistics at the claim level (N = 16,212,409 claims). Distance traveled is measured in miles. Charges are reported in dollars.

Facility Level Data

To supplement the claims data, I incorporate facility-level information from multiple external sources. First, I utilize data supplied by the New York State Department of Health, which provides detailed information on each facility's geographic location and ownership structure. The ownership data allow me to identify facilities affiliated with the same hospital system. Furthermore, I acquire ACO level data from the Centers for Medicare & Medicaid Services (CMS), which provides details which ACO each facility belongs to, if any. Furthermore, CMS provides data on the benchmark spending level of each ACO, and the share of ACO performance retained by the government. Table 2 provides facility level summary statistics.

In the sample period, 42% of facilities participates in an ACO, with an average of 6 facilities per ACO. ACOs receive an annual per patient benchmark of $13,014. In the sample period, 51% of facilities are part of a multi-facility system with an average of 4.68 facilities per system.

2.5 Measures For Structural Model

In analyzing the effect of ACOs on spending and quality, several empirical challenges arise due to the lack of direct facility-level measures. In the model, systems choose quantity of

Table 2: Facility Sample Summary Statistics

	Mean	Std. Dev.
Dummy Variable For ACO participation	0.42	
Number of Facilities Per ACO	6.00	4.27
Per-Patient Benchmark	13,014	1,499
Dummy Variable For Multi-Facility System	0.51	
Number of Facilities Per System	4.68	3.38

Notes: This table includes 678 facility-year level observations. Statistics for facilities per ACO are conditional on participating in an ACO. Statistics for facilities per system are conditional on the the system being multi-facility.

services and quality of care to maximize profit. Neither the quantity of services provided nor the quality of care delivered is directly observed in the data, so both must be constructed from claim-level information.

Quantity of Services To measure the average quantity of services at the facility-year level, I first divide each patient's outpatient spending by the reimbursement rate for their diagnosis, obtaining a patient-level measure of service quantity. I then average these patient-level quantities across all patients treated at the facility in that year, following Ho and Lee (2017). See Appendix A.3 for details.

Quality of Care Following the literature on measuring quality for healthcare providers (Gowrisankaran and Town, 1999; Geweke et al., 2003; Gaynor et al., 2016), I use health outcomes of patients that visit a facility in a given year to measure the quality of each facility-year. The health outcome I use is transition from outpatient care to inpatient care within 30 days of visiting an outpatient facility. This quality measure based on health outcomes is most similar to Einav et al. (2025), who uses transitions from Skilled Nursing Facilities to inpatient care within 90 days. Other studies in the literature have examined similar quality metrics, such as mortality within 14 days, as in Gowrisankaran and Town (1999), Geweke et al. (2003), Gaynor et al. (2016). I focus on transition from outpatient care to inpatient care for, as opposed to the more commonly examined mortality measure for two reasons. First, mortality rarely occurs at the outpatient level. In my sample, mortality at the outpatient level occurred in less than 0.01% of cases. Second, transitions from outpatient to inpatient care is the quality metric most targeted by facilities that participate in an ACO (CMS, 2025a), as reductions in transitions from outpatient care to inpatient care improves ACO performance through lower inpatient spending.

To measure quality of care at the facility-year level, I first construct a dummy variable at the patient-facility-diagnosis-year level that takes on value 1 if the patient is observed transitioning from outpatient care to inpatient care within 30 days of visiting the outpatient facility. Next, I regress this dummy variable on a set of facility-year fixed effects and controls at the individual and diagnosis level. The facility-year fixed effects are the main measure for quality at the facility-year level, where the individual and diagnosis level controls take into account that different diagnoses and individuals with different underlying health transition from outpatient care to inpatient care at different rates. Once the parameters of this regression are estimated, I use the fitted value of this regression, evaluated at the average levels of patient-level and diagnosis-level controls, as the empirical measure of quality for each facility-year. See Appendix A.4 for details.

Patient Assignment Patient assignment to ACOs depends on the patient's level of primary care spending. In my model, I convert observable outpatient spending to primary care spending, using a conversion factor that varies by diagnosis category. To calculate this conversion factor from the data, I take the average ratio between outpatient spending on primary care and overall outpatient spending, averaged over all claims within a diagnosis category.

Statisitics of Endogenous Variables Table 3 provides summary statistics for the key endogenous variables of the model

Table 3: Summary Statistics - Endogenous Variables

	Mean	Std. Dev.	Q25	Median	Q75
Quantity of services	642	213	486	621	796
Quality of care	0.96	0.03	0.94	0.95	0.98

Notes: This table includes 678 facility-year level observations. See text for details on constructing measures for endogenous variables. Quantity of services is measured in units of services. Quality of care is defined as one minus the proportion of patients that transition from outpatient to inpatient within 30 days of visiting the outpatient facility.

The mean quantity of services provided is 642, which is measured in units of service. On average, this translates to a payment of $1,798 per visit once multiplied by the reimbursement rate. The mean quality of care is 0.96, which can be interpreted as the mean transition rate from outpatient to inpatient care across facilities is 4% $(1 - 0.96)$.

3 Model

In this section, I construct a structural model of demand and supply in the Medicare outpatient facility services market. The model consists of two stages. In the first stage, multi-facility hospital systems simultaneously choose quantity of services and quality of care for all facilities within the system to maximize profit.[8] Facility choices of quantity of services map onto a per-patient payment. Systems who participate in an ACO are incentivized by the performance of the ACO where system participation in ACOs are exogenously given. In the second stage, patients who are enrolled in either Traditional Medicare or Medicare Advantage observe the choices of the first stage of the game. For each realized diagnosis, each patient chooses a facility that maximizes her utility. Once firm and patient choices are realized, each patient is assigned to either one ACO or left unassigned, determined by realized spending.[9]

3.1 Demand

Patients make a discrete choice among differentiated outpatient facilities. In year t, patient i with diagnosis category d covered by insurer type m chooses facility j from her choice set $\mathcal{CS}_{imd}$ that maximizes her utility.[10] Each facility is characterized by a spending level p_{jd}, quality of care q_j, and a matrix of observable patient-facility characteristics z_{ijmd}. The spending of facility j for diagnosis category d is a function of facility-specific quantity of services x_j and a diagnosis category specific reimbursement rate r_d, where r_d is measured in dollars per unit of quantity of services. The utility consumer i receives from facility j is

$$u_{ijmd} = \alpha_m p_{jd}(r_d, x_j) + \beta q_j + z_{ijmd}\gamma + \xi_{jmd} + \epsilon_{ijmd} \tag{2}$$

where α_m captures insurer-type-specific preferences for spending, β captures preferences for quality and γ is a vector of preference parameters for characteristics that vary at the patient-facility level.[11] Consumers have additional preferences for facilities that vary at the facility-insurer-diagnosis-year level ξ_{jmd} and idiosyncratic preferences ϵ_{ijmd}, both of which

[8] Quality of care is defined as one minus the proportion of patients that transition from outpatient to inpatient within 30 days of visiting the outpatient facility. Under this definition, higher values of the quality measure are interpreted as "higher quality" facilities.

[9] See section 2.3 for more details on patient assignment.

[10] For exposition, I suppress the year index t.

[11] One could allow α_m to vary with individual characteristics as well as insurer categories. For simplicity, however, I consider heterogeneity in price sensitivity only across insurer categories, which captures differences arising from the varying generosity (expected coinsurance rates and copayments) of each insurer type.

are observed by the patient but not by the econometrician.

The patient also has the option of choosing a non-hospital-based provider. The utility from the outside option, indexed as $j = 0$, is

$$u_{i0md} = \alpha_m p_{0d}(r_d, x_0) + \beta q_0 + \epsilon_{i0md}. \tag{3}$$

Patient i chooses facility j if

$$h_{ijmd} = \mathbf{1}(u_{ijmd} > \max_{j' \neq j \in \mathcal{CS}_{imd}} u_{ij'md}) \tag{4}$$

where h_{ijmd} is a dummy variable that takes value 1 if patient i, covered by insurer type m with diagnosis category d, chooses facility j.

Assuming that ϵ_{ijmd} is distributed i.i.d. Type 1 Extreme Value, the probability that patient i chooses facility j is

$$s_{ijmd} = \Pr(h_{ijmd} = 1) = \frac{exp\{\alpha_m p_{jd}(r_d, x_j) + \beta q_j + z_{ijmd}\gamma + \xi_{jmd}\}}{\sum_{j' \in \mathcal{CS}_{imd}} exp\{\alpha_m p_{j'd}(r_d, x_{j'}) + \beta q_{j'} + z_{ij'md}\gamma + \xi_{j'mt}\}}. \tag{5}$$

The demand for facility j from patients who are enrolled in Traditional Medicare is

$$D_j^{\text{TM}}(\boldsymbol{x}, \boldsymbol{q}) = \sum_{i \in \mathcal{I}^{\text{TM}}} s_{ijmd}(\boldsymbol{x}, \boldsymbol{q}) \tag{6}$$

where $\boldsymbol{x}, \boldsymbol{q}$ are vectors of each facilities' choices of quantity of services and quality of care, and $\mathcal{I}^{\text{TM}}$ is the set of patients enrolled in Traditional Medicare.

Similarly, the demand for facility j from patients who are enrolled in Medicare Advantage is

$$D_j^{\text{MA}}(\boldsymbol{x}, \boldsymbol{q}) = \sum_{i \in \mathcal{I}^{\text{MA}}} s_{ijmd}(\boldsymbol{x}, \boldsymbol{q}) \tag{7}$$

where $\mathcal{I}^{\text{MA}}$ is the set of patients enrolled in Medicare Advantage.

3.2 Supply

I consider a full information static oligopoly game where multi-facility hospital systems compete in average quantity of services and quality of care in a Nash-Bertrand fashion. I define the market as a year-geographic region (New York State).[12]

Let $\mathcal{J}_f$ denote the set of facilities in system f. In a given year, each hospital system f simultaneously chooses an average level of quantity of services x_j and quality of care q_j for all facilities in system f to maximize

$$\max_{\{x_j, q_j\}_{j \in \mathcal{J}_f}} \Pi_f^k(\boldsymbol{x}, \boldsymbol{q}) \tag{8}$$

where Π_f^k is the profit of system f participating in ACO k. Each facility's level of quantity of services is mapped to a per-visit payment $p_j(x_j)$. Hospital systems participate in a chosen ACO k, where $k = 0$ indexes not participating in an ACO.[13] The profit of hospital system f that participates in ACO k is

$$\Pi_f^k = \sum_{j \in \mathcal{J}_f} \left[p_j(x_j) - mc_j(q_j) \right] D_j(\boldsymbol{x}, \boldsymbol{q}) + w_f \pi_k(\boldsymbol{x}, \boldsymbol{q}) \tag{9}$$

where $D_j(\boldsymbol{x}, \boldsymbol{q}) = D_j^{\text{TM}}(\boldsymbol{x}, \boldsymbol{q}) + D_j^{\text{MA}}(\boldsymbol{x}, \boldsymbol{q})$ is the total demand for facility j across Traditional Medicare patients and Medicare Advantage patients, $mc_j(q_j)$ is the marginal cost of treating one patient, which varies by quality, and $\pi_k(\boldsymbol{x}, \boldsymbol{q})$ is the performance of ACO k, which is a function of all facility choices of quantity of services and quality of care, including facilities that do not participate in ACO k.[14] The performance of the ACO is split across all systems in the ACO according to the sharing rule $w_f \in [0, 1]$. I assume that each facility commits to one level of quantity of services and quality of care for each Medicare patient, regardless of whether the patient is enrolled in Traditional Medicare and Medicare Advantage. Similarly, I assume the marginal cost to treat one patient does not depend on whether the patient is enrolled in Traditional Medicare and Medicare Advantage.

I specify the marginal cost function as

[12] Consistent with the demand model, I suppress the year index t in the supply model.

[13] Systems' choices of which ACO to join are not explicitly modeled and are treated as given.

[14] I hold $mc_j(q_j)$ constant in x_j, as the optimality conditions do not separately identify the marginal cost level and its slope with respect to x_j.

$$mc_j(q_j) = c_j^o + \frac{\partial mc_j}{\partial q_j} q_j \tag{10}$$

where c_j^o captures components of marginal cost that do not vary by quality and $\frac{\partial mc_j}{\partial q_j}$ captures the slope of the marginal cost function with respect to quality.

The performance of ACO k can be decomposed expressed as the ACO's benchmark, outpatient spending and inpatient spending. The performance of ACO k is

$$\begin{aligned} \pi_k(\boldsymbol{x}, \boldsymbol{q}) = & \underbrace{b_k \sum_j \sum_{i \in \mathcal{I}^{\text{TM}}} s_{ijmd}(\boldsymbol{x}, \boldsymbol{q}) a_{ik}(\boldsymbol{x})}_{\text{ACO's Benchmark}} - \underbrace{\sum_j p_j(x_j) \sum_{i \in \mathcal{I}^{\text{TM}}} s_{ijmd}(\boldsymbol{x}, \boldsymbol{q}) a_{ik}(\boldsymbol{x})}_{\text{ACO's Outpatient Spending}} \\ & - \underbrace{\sum_j (1 - q_j) Y_j \sum_{i \in \mathcal{I}^{\text{TM}}} s_{ijmd}(\boldsymbol{x}, \boldsymbol{q}) a_{ik}(\boldsymbol{x})}_{\text{ACO's Inpatient Spending}} \end{aligned} \tag{11}$$

where b_k is the benchmark per assigned patient of ACO k, a_{ik} is the probability that patient i is assigned to ACO k conditional on realized choices of facilities, s_{ijmd} is the probability that patient i chooses facility j and Y_j is the inpatient spending per consumer for facility j.[15] The performance of the ACO compares the difference between the ACO's benchmark and the sum of outpatient and inpatient spending of patients assigned to the ACO. The term $s_{ijmd}(\boldsymbol{x}, \boldsymbol{q}) a_{ik}(\boldsymbol{x})$ gives the joint probability of patient i choosing facility j and is assigned to ACO k. Summing up this joint probability over all patients $i \in \mathcal{I}^{\text{TM}}$ yields the number of patients that choose facility j and are assigned to ACO k.

ACO's Benchmark The benchmark of ACO k is given by the product of the per-patient benchmark b_k and the total number of patients assigned to ACO k. Summing up the number of patients that choose facility j and are assigned to ACO k over all facilities j in the market yields the total expected number of patients assigned to ACO k. The per-patient benchmark b_k is calculated as the average Medicare spending across all patients in the counties served by ACO k. In practice, this benchmark is risk-adjusted for differences in patient health severity. Due to data limitations, I do not apply individual-level risk adjustment; instead, I use an aggregated per-patient benchmark that accounts for the case mix of each ACO.

[15] I assume that Y_j is exogenous, which is justified by the fact that Medicare exogenously sets inpatient reimbursement rates, following the Diagnosis Related Group (DRG) reimbursement system, which facilities take as given.

ACO's Outpatient Spending The ACO's outpatient spending can be found by multiplying the number of patients who choose facility j and are assigned to ACO k, $\sum_{i \in \mathcal{I}^{\text{TM}}} s_{ijmd}(\boldsymbol{x}, \boldsymbol{q}) a_{ik}(\boldsymbol{x})$, by the per-patient spending at facility j, $p_j(x_j)$, and summing up over all facilities in the market. Notice that this summation over facilities includes facilities who participate in the ACO and facilities that do not. This plays an important role in the incentives of the ACO; if a large number of patients that are assigned to the ACO choose a facility j' with high outpatient spending ($p_{j'}(x_{j'})$ is high), the ACO's performance will decrease, regardless of whether or not the high spending outpatient facility is in the ACO. On the contrary, if the assigned patients choose a facility with low outpatient spending, the ACO's performance will be better relative to the case of a large number of patients choosing a high spending facility.

ACO's Inpatient Spending The ACO's inpatient spending can be found by starting with the number of patients who choose facility j and are assigned to ACO k, $\sum_{i \in \mathcal{I}^{\text{TM}}} s_{ijmd}(\boldsymbol{x}, \boldsymbol{q}) a_{ik}(\boldsymbol{x})$; multiplying this term by Y_j gives the inpatient spending of patients who choose facility j and are assigned to ACO k, conditional on requiring inpatient care. Multiplying this quantity by $(1 - q_j)$ yields the unconditional expected inpatient spending of these patients. Finally, summing over all facilities j gives the total expected inpatient spending of patients assigned to ACO k. Notice that this includes facilities who participate in the ACO and facilities that do not. If a large number of patients that are assigned to the ACO choose a facility j' with low quality ($q_{j'}$ is low), this will worsen the ACO's performance through higher inpatient spending. On the contrary, if many patients assigned to the ACO choose a high quality facility, the ACO's performance will be better relative to the case of a large number of patients choosing a low quality facility. By ACO performance depending partially on inpatient spending for patients assigned to the ACO, participants in an ACO now directly internalize the quality of facilities in the market beyond traditional demand effects.[16]

Penalties From ACO Performance An ACO may have negative performance, $\pi_k < 0$, if the ACO's benchmark is less than the total spending of the ACO. In this case, the participants of the ACO must pay a penalty to the government. For participation in an ACO to be potentially profitable (that is, $\pi_k > 0$), the per patient benchmark b_k needs to be sufficiently high. More specifically, the per-patient benchmark of the ACO needs to exceed the expected total of outpatient and inpatient spending for each patient. Otherwise, the ACO's penalties from outpatient and inpatient spending will outweigh the benefits from the benchmark.

[16] "Traditional demand effects" refers to the incentive for a facility to increase quality in order to increase demand.

Assignment to ACOs To complete the supply model, I discuss how patients' assignment probabilities to ACOs are constructed. After patient choices of facilities are realized, patients are assigned to one ACO or are not assigned to any ACO.[17] The total primary care spending for patient i incurred by facilities in ACO k is

$$v_{ik} = \sum_{d} PCP_d \times g_{idk}(\boldsymbol{x}_k) + e_{ik} \tag{12}$$

where v_{ik} is the primary care spending incurred by patient i under facilities in ACO k, $g_{idk}(\boldsymbol{x}_k)$ is the observed outpatient spending patient i accumulated for diagnosis category d under facilities in ACO k, $\boldsymbol{x}_k$ is a vector of quantity of services of all facilities in ACO k, PCP_d is the fraction of spending that is classified as primary care spending for diagnosis category d which is treated as data, and e_{ik} is unobserved primary care spending accumulated by patient i under facilities that participate in ACO k.[18] Because primary care services and associated primary care spending make up only a subset of outpatient services, PCP_d is a value between 0 and 1.

Patient i is assigned to ACO k if

$$A_{ik} = \mathbf{1}(v_{ik} > \max_{k' \neq k} v_{ik'}). \tag{13}$$

Let $\boldsymbol{h}_i$ be the vector of realizations for all facility choices for patient i. Assuming that e_{ik} is distributed i.i.d. Type 1 Extreme Value, the probability that patient i is assigned to ACO k, conditional on patient choices over facilities, is

$$a_{ik}(\boldsymbol{x}) = \Pr(A_{ik} = 1 | \boldsymbol{h}_i) = \frac{exp\{\sum_d PCP_d \times g_{idk}(\boldsymbol{x}_k)\}}{\sum_{k' \in \mathcal{K}} exp\{\sum_d PCP_d \times g_{idk'}(\boldsymbol{x}_{k'})\}} \tag{14}$$

where $\mathcal{K}$ is the set of all ACOs in the market. An individual may be assigned to no ACO, $k = 0$, whose probability takes the same functional form as being assigned to an ACO.

[17] Unlike the demand model for facilities characterized in the previous section, patients do not make an active decision in which ACO they are assigned. Assignment depends on the facility choice stage outlined in section 3.1 and spending conditional on facility choices.

[18] Unobserved primary care spending may exist in the data, due to claims that happen in a facility outside of New York State, which would not be observed in the data.

First Order Conditions

Next, I derive the systems' first order conditions. The first order condition for facility j with respect to quantity of services x_j is

$$\frac{\partial p_j}{\partial x_j} D_j + \sum_{j' \in \mathcal{J}_f} \frac{\partial D_{j'}}{\partial x_j} \left[p_{j'}(x_{j'}) - mc_{j'}(q_{j'}) \right] + w_f \frac{\partial \pi_k}{\partial x_j} = 0, \tag{15}$$

and the first order condition for facility j with respect to quality of care q_j is

$$-\frac{\partial mc_j}{\partial q_j} D_j + \sum_{j' \in \mathcal{J}_f} \frac{\partial D_{j'}}{\partial q_j} \left[p_{j'}(x_{j'}) - mc_{j'}(q_{j'}) \right] + w_f \frac{\partial \pi_k}{\partial q_j} = 0. \tag{16}$$

Moving to how marginally increasing the quantity of services affects ACO performance, the partial derivative of ACO performance with respect to quantity of services is

$$\begin{aligned}\frac{\partial \pi_k}{\partial x_j} &= b_k \big(\sum_{j'} \sum_{i \in \mathcal{I}^{\text{TM}}} \frac{\partial s_{ij'md}}{\partial x_j} a_{ik} + s_{ij'md} \frac{\partial a_{ik}}{\partial x_j} \big) - \frac{\partial p_j}{\partial x_j} \sum_{i \in \mathcal{I}^{\text{TM}}} s_{ijmd} a_{ik} \\ &\quad - \sum_{j'} p_{j'} \big(\sum_{i \in \mathcal{I}^{\text{TM}}} \frac{\partial s_{ij'md}}{\partial x_j} a_{ik} + s_{ij'md} \frac{\partial a_{ik}}{\partial x_j} \big) - \sum_{j'} (1 - q_{j'}) Y_{j'} \big(\sum_{i \in \mathcal{I}^{\text{TM}}} \frac{\partial s_{ij'md}}{\partial x_j} a_{ik} + s_{ij'md} \frac{\partial a_{ik}}{\partial x_j} \big).\end{aligned} \tag{17}$$

Similarly, the partial derivative of ACO performance with respect to quality of care is

$$\begin{aligned}\frac{\partial \pi_k}{\partial q_j} &= b_k \big(\sum_{j'} \sum_{i \in \mathcal{I}^{\text{TM}}} \frac{\partial s_{ij'md}}{\partial q_j} a_{ik} + s_{ij'md} \frac{\partial a_{ik}}{\partial q_j} \big) - \sum_{j'} p_{j'} \big(\sum_{i \in \mathcal{I}^{\text{TM}}} \frac{\partial s_{ij'md}}{\partial q_j} a_{ik} + s_{ij'md} \frac{\partial a_{ik}}{\partial q_j} \big) \\ &\quad + Y_j \sum_{i \in \mathcal{I}^{\text{TM}}} s_{ijmd} a_{ik} - \sum_{j'} (1 - q_{j'}) Y_{j'} \big(\sum_{i \in \mathcal{I}^{\text{TM}}} \frac{\partial s_{ij'md}}{\partial q_j} a_{ik} + s_{ij'md} \frac{\partial a_{ik}}{\partial q_j} \big).\end{aligned} \tag{18}$$

Incentives of ACOs

Prior to discussing the equilibrium impacts of the incentives introduced by ACOs, I first discuss the incentives and mechanisms introduced by ACOs. First, consider the first order conditions with respect to quantity of services derived in equation 17. The incentives introduced by ACOs in this equation have four effects:

(*i*) *Benchmark effect*: The term $b_k \big(\sum_{j'} \sum_{i \in \mathcal{I}^{\text{TM}}} \frac{\partial s_{ij'md}}{\partial x_j} a_{ik} + s_{ij'md} \frac{\partial a_{ik}}{\partial x_j} \big)$ captures the incentive associated with changing the benchmark of the ACO; if facility j increases its quantity of services, this leads to a change in the number of patients assigned to the ACO, which in turn changes the benchmark of the ACO.

(ii) *Own outpatient spending effect*: The term $-\frac{\partial p_j}{\partial x_j}\sum_{i\in\mathcal{I}^{\text{TM}}} s_{ijmd}a_{ik}$ captures the impacts of lower ACO performance through higher spending accumulated by facility j. If facility j increases its quantity of services, the total outpatient spending of patients assigned to the ACO increases, which worsens ACO performance.

(iii) *Outpatient spending substitution effect*: The term $-\sum_{j'} p_{j'}\big(\sum_{i\in\mathcal{I}^{\text{TM}}}\frac{\partial s_{ij'md}}{\partial x_j}a_{ik}+s_{ij'md}\frac{\partial a_{ik}}{\partial x_j}\big)$ captures the internalization of other facilities' spending on ACO performance. If facility j increases its quantity of services, this affects the demand of facility j and other facilities $j'\neq j$. One can interpret this as the firm internalizing the substitutability between facilities; if facility j increases its quantity of services, patients may substitute towards other facilities who may have high (low) outpatient spending which decreases (increases) ACO performance.[19]

(iv) *Inpatient spending substitution effect*: The term $-\sum_{j'}(1-q_{j'})Y_{j'}\big(\sum_{i\in\mathcal{I}^{\text{TM}}}\frac{\partial s_{ij'md}}{\partial x_j}a_{ik}+s_{ij'md}\frac{\partial a_{ik}}{\partial x_j}\big)$ captures the substitution effect associated with inpatient spending. If facility j increases its quantity of services, patients may substitute towards other facilities who may have high (low) quality of care, which increases (decreases) ACO performance through lower (higher) inpatient spending.

Next, I turn to the partial derivative of ACO performance with respect to quality of care derived in equation 18. The incentives introduced by ACOs have four effects:

(i) *Benchmark effect*: The term $b_k\big(\sum_{j'}\sum_{i\in\mathcal{I}^{\text{TM}}}\frac{\partial s_{ij'md}}{\partial q_j}a_{ik}+s_{ij'md}\frac{\partial a_{ik}}{\partial q_j}\big)$ captures the effect of quality on the ACO's benchmark. If the facility increases its quality, this changes the number of patients assigned to the ACO, which changes the benchmark of the ACO.

(ii) *Outpatient spending substitution effect*: The term $-\sum_{j'} p_{j'}\big(\sum_{i\in\mathcal{I}^{\text{TM}}}\frac{\partial s_{ij'md}}{\partial q_j}a_{ik}+s_{ij'md}\frac{\partial a_{ik}}{\partial q_j}\big)$ captures the substitution effect induced by a change in quality; an increase in the qual-

[19]This incentive is similar to the cannibalization effect commonly discussed in the multi-product competition literature in industrial organization (Nevo, 2001). In such settings, a price increase on one product may shift consumers toward other products the firm also owns, which the firm internalizes when making pricing decisions. Similarly, in my setting, the participant in the ACO internalizes the effect of consumers switching to other facilities, induced by an increase in the quantity of services, through its impact on ACO performance. Since ACO performance depends on spending across all facilities, this cannibalization effect extends to all facilities in the market, including facilities outside the ACO. The strength of the effect depends on the cross-substitutability between facilities.

ity of facility j affects the demand of facility j and other facilities $j' \neq j$, which changes ACO performance through a change in outpatient spending.

(*iii*) *Own inpatient spending effect*: The term $Y_j \sum_{i \in \mathcal{I}^{\text{TM}}} s_{ijmd} a_{ik}$ captures the effect of an increase in the quality of facility j on inpatient spending; if facility j improves its quality of care, the inpatient spending incurred by patients assigned to the ACO that choose facility j decreases, which improves ACO performance.

(*iv*) *Inpatient substitution effect*: The term $-\sum_{j'}(1-q_{j'})Y_{j'}\big(\sum_{i \in \mathcal{I}^{\text{TM}}} \frac{\partial s_{ij'md}}{\partial q_j} a_{ik} + s_{ij'md}\frac{\partial a_{ik}}{\partial q_j}\big)$ captures a substitution effect present in inpatient spending. If facility j increases its quality of care, this changes the demand of facility j and other facilities $j' \neq j$, where patients may be steered away from other facilities who have high (low) quality of care, which increases (decreases) ACO performance through lower (higher) inpatient spending.

3.3 Equilibrium

In each market (year), the Nash-Bertrand game generates a set of $2 \times J$ first order conditions, where J is the number of facilities in the market.

I first define the $J \times J$ matrix of own and cross derivatives of demand with respect to quantity of services, accounting for system ownership of facilities, as

$$\Delta^x(j,j') = \begin{cases} \sum_i \frac{\partial s_{ij'md}}{\partial x_j} \text{ if } j', j \in \mathcal{J}_f \\ 0 \text{ otherwise.} \end{cases} \tag{19}$$

Similarly, I define the $J \times J$ matrix containing facilities' own and cross derivatives of demand with respect to quality of care as

$$\Delta^q(j,j') = \begin{cases} \sum_i \frac{\partial s_{ij'md}}{\partial q_j} \text{ if } j', j \in \mathcal{J}_f \\ 0 \text{ otherwise.} \end{cases} \tag{20}$$

Next, I define $\boldsymbol{D}$ as a $J \times 1$ vector of facility demand, where $[\boldsymbol{D}]_j = D_j^{\text{TM}} + D_j^{\text{MA}}$. Similarly, I define the $J \times 1$ vector of facility payments as $\boldsymbol{p(x)}$, where $[\boldsymbol{p(x)}]_j = p_j(x_j)$. In addition, I define $\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{x}}$ as a $J \times 1$ vector of partial derivatives of payments with respect to quantity of services, where $[\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{x}}]_j = \frac{\partial p_j}{\partial x_j}$. Moving onto marginal cost terms, I define $\boldsymbol{mc}$ as the $J \times 1$ vector of stacked marginal cost terms, where $[\boldsymbol{mc}]_j = mc_j$. The $J \times 1$ vector of partial derivatives of marginal cost with respect to quality is $\frac{\partial \boldsymbol{mc}}{\partial \boldsymbol{q}}$, where $[\frac{\partial \boldsymbol{mc}}{\partial \boldsymbol{q}}]_j = \frac{\partial mc_j}{\partial q_j}$. Moving onto terms involving ACO performance, I define $\boldsymbol{w}$ as the $J \times 1$ vector of sharing rules, where $[\boldsymbol{w}]_j = w_{f(j)}$. Finally, I define $\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{x}}$ and $\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{q}}$ as $J \times 1$ vectors of the partial derivative of ACO performance with respect to quantity of services and the partial derivative of ACO performance with respect to quality of care where $[\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{x}}]_j = \frac{\partial \pi_{k(j)}}{\partial x_j}$ and $[\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{q}}]_j = \frac{\partial \pi_{k(j)}}{\partial q_j}$.

Stacking all first order conditions with respect to quantity of services across all providers in year t yields the system of equations,

$$diag(\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{x}})\boldsymbol{D} + \boldsymbol{\Delta}^{\boldsymbol{x}}(\boldsymbol{p}(\boldsymbol{x}) - \boldsymbol{mc}) + diag(\boldsymbol{w})\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{x}} = 0, \tag{21}$$

where the $diag(\cdot)$ operator inputs a vector of size $J \times 1$ and outputs a $J \times J$ matrix with the elements of the input on the diagonal and zeros on the off-diagonal. Likewise, stacking the quality of care first order conditions across all facilities in the market yields

$$-diag(\boldsymbol{D})\frac{\partial \boldsymbol{mc}}{\partial \boldsymbol{q}} + \boldsymbol{\Delta}^{\boldsymbol{q}}(\boldsymbol{p}(\boldsymbol{x}) - \boldsymbol{mc}) + diag(\boldsymbol{w})\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{q}} = 0. \tag{22}$$

The equilibrium vector of quantity of services $\boldsymbol{x}^*$ and quality $\boldsymbol{q}^*$ for a market simultaneously satisfies the set of first order conditions in equations 21 and 22.[20]

3.4 Effects of ACO Incentives

In this section, I discuss the effects of ACO incentives on equilibrium outcomes. In addition to the direct effects of ACOs, I highlight the spillover effects on systems that do not participate in an ACO and spillover effects onto patients that are not counted toward ACO performance.

Effects of ACO Incentives on Participants

The effects of the incentives introduced by ACOs on equilibrium outcomes of quantity of services and quality of care on participants of an ACO are ambiguous. The equilibrium effects depend on the magnitude and signs of $\frac{\partial \pi_k}{\partial x_j}$ and $\frac{\partial \pi_k}{\partial q_j}$, the magnitude of the sharing rule w_f, the underlying demand system and the marginal cost curve of each facility.

Next, I characterize the potential equilibrium outcomes driven by ACO incentives for ACO participants. Abstracting away from multi-facility hospital systems, competition and

[20]Because the strategy spaces for facilities are nonempty compact convex subsets of a Euclidean space and the profit function is concave in its own strategies, there exists a pure-strategy Nash equilibrium following Fudenberg and Tirole (1991). There may be multiple equilibria in the game. Exogenous spatial heterogeneity and heterogeneity in facility costs can make multiple equilibria less likely, but by itself does not ensure uniqueness.

cannibalization effects for exposition, consider the case of a simple single hospital system. The first order condition for the single facility system with respect to quantity of services is

$$\frac{\partial p_j}{\partial x_j} D_j + \frac{\partial D_j}{\partial x_j}\big(p_j(x_j) - mc_j(q_j)\big) + w_f \frac{\partial \pi_k}{\partial x_j} = 0. \tag{23}$$

Notice that the usual profit maximizing incentives are present in these first order conditions: increasing x_j leads to higher revenue at the cost of decreasing demand. Suppose $\frac{\partial \pi_k}{\partial x_j}$ is negative; that is, increasing quantity of services leads to lower ACO performance. This would decrease the marginal benefit of increasing quantity of services, which may lower the equilibrium value of x_j^*.

Now consider the first order condition with respect to quality for the single facility system

$$-\frac{\partial mc_j}{\partial q_j} D_j + \frac{\partial D_j}{\partial q_j}\big(p_j(x_j) - mc_j(q_j)\big) + w_f \frac{\partial \pi_k}{\partial q_j} = 0. \tag{24}$$

Notice again that the usual profit maximizing incentives are present: increasing quality of results in the benefit of higher demand for the facility at the cost of a higher marginal cost. Suppose $\frac{\partial \pi_k}{\partial q_j}$ is positive; that is, increasing quality of care leads to higher ACO performance. This would increase the marginal benefit of increasing quality, potentially leading to a higher equilibrium value of q_j^*.

While the incentives of ACOs are clear at the margin, the overall equilibrium outcomes are rendered more complex when simultaneously considering decisions with respect to quantity of services and quality of care. Consider the term $\frac{\partial \pi_k}{\partial x_j} < 0$ in equation 23, which lowers the marginal benefit of increasing quantity of services for ACO participants. This may induce a reduction in outpatient spending; however, this will lower the marginal benefit to increasing quality captured in the term $\frac{\partial D_j}{\partial q_j}\big(p_j(x_j) - mc_j(q_j)\big)$ in equation 24, even under the presence of a higher marginal benefit to increasing quality captured by the term $\frac{\partial \pi_k}{\partial q_j} > 0$ in equation 24. In this scenario, the facility will have a lower equilibrium value of x_j^* and lower equilibrium value of q_j^*, relative to a regime without ACO incentives, which may be a case of quality underprovision with lower spending. Whether or not this trade-off between lower spending and lower quality leads to higher or lower consumer welfare depends on consumer preferences over spending and quality.

Next, consider a case where the marginal incentives of ACOs increases the quality of the facility, driven by the term $\frac{\partial \pi_k}{\partial q_j} > 0$ in equation 24. This increased quality may results in higher costs for the facility, captured by a higher value of $mc_j(q_j)$. Turning to the response in quantity of services, the higher marginal costs causes a lower marginal disutility of an increase in quantity of services, captured by the term $\frac{\partial D_j}{\partial q_j}\big(p_j(x_j) - mc_j(q_j)\big)$ in equation 23. This decreased marginal distuility to increasing quantity of services, may induce a higher equilibrium value of x_j^*, even under the presence of the increased marginal disutility to increasing quantity of services caused by ACO incentives, which is captured by $\frac{\partial \pi_k}{\partial x_j} < 0$ in equation 23. In this scenario, the facility will have a higher equilibrium value of x_j^* and a higher equilibrium value of q_j^*, relative to a regime without ACO incentives, which may be a case of quality overprovision. Similar to the quality underprovision case, the qualitative direction of consumer welfare in the case of a trade-off between higher spending and higher quality depends on consumer preferences over spending and quality (Crawford et al., 2019).

Finally, consider a case where the incentives of ACOs are strong enough to ensure an equilibrium where x_j^* decreases and q_j^* increases, relative to a regime without ACO incentives. In other words, the incentives of ACOs do not create a trade-off between spending and quality, as characterized in the two previous cases. This outcome occurs when ACO incentives are strong enough to offset the trade-offs between spending and quality, either preventing quality disinvestment when spending decreases or limiting spending increases that would otherwise be used to cover the costs of higher quality. Abstracting away from competitive responses by systems that do not participate in an ACO, the lower spending and higher quality created by ACO incentives increases consumer welfare.

Spillover Effects

In the model, systems choose the same levels of quantity of services and quality of care for both Traditional Medicare and Medicare Advantage patients. ACO incentives shift the equilibrium values of quantity of services and quality of care, generating spillover effects onto Medicare Advantage patients, even though they are not directly targeted by the program. Ignoring Medicare Advantage patients overstates the equilibrium impacts of ACOs by failing to account for their demand. It also biases the estimated welfare effects, as it omits the contributions of Medicare Advantage patients to overall consumer welfare.

Impact on Welfare

ACO incentives can create a trade-off between spending and quality, making their welfare effects ambiguous. The resulting equilibrium may involve quality overprovision with higher spending, quality underprovision with lower spending or a desirable outcome where spending decreases and quality improves. Ultimately, the welfare impact of ACOs depends on how competing systems respond in equilibrium and on consumer preferences. To evaluate the equilibrium and welfare impacts of ACOs, I first estimate the parameters of the demand system and recover each facility's marginal cost curve. Using these recovered primitives, I then simulate counterfactual scenarios to quantify the welfare effects of ACOs.

4 Estimation

In this section, I discuss the empirical specification and estimation strategy for the demand and supply model. I estimate the demand and supply parameters separately. First, I estimate the demand parameters. Then, I take the demand estimates as given and estimate the parameters of the supply model.

4.1 Demand Estimation

Empirical Specification

I consider four insurer categories for the price sensitivity parameter α_m: Traditional Medicare and Medicare Advantage, each with or without Medicaid coverage. Consumers who enroll in Traditional Medicare have different price sensitivities than consumers who enroll in Medicare Advantage, due to different generosities across insurance coverage. Consumers with Medicaid in addition to either Traditional Medicare or Medicare Advantage have all medical costs covered. To this end, I set α_m for consumers who have Medicaid coverage equal to 0.[21]

Included in the individual-facility level variables z_{ijmd} are a first and second order distance term, a dummy variable for past use of the facility, a dummy variable for past use of a facility within the same system and a dummy variable for past use of a facility within the same ACO. The first order distance term is included to allow for disutility from travel distance, while the second order term allows for decreasing marginal disutility from far away facilites. The dummy variable for past use of a facility is included to capture inertia of re-visiting a facility for a subsequent outpatient visit. I include the dummy variables for the past use of a facility within the same system and the past use of a facility within the same ACO to allow for steering behavior within a system or ACO.[22]

[21]Some patients who are enrolled in Traditional Medicare may have additional supplementary insurance besides Medigap; for example, Medigap or employer sponsored supplementary insurance. Similar to consumers with Medicaid, these consumers face little to no out-of-pocket costs. Unfortunately, I do not observe if a patient enrolled in Traditional Medicare has supplementary insurance beyond Medicaid. Therefore, my estimate of spending sensitivity for Traditional Medicare patients is the mean price sensitivity across patients with and without supplementary insurance.

[22]Steering behavior within a system or ACO can be interpreted as referral behavior. Unfortunately, I do not observe referral decisions in my data. Instead, I proxy for referrals using dummy variables for past use of a facility within the same system and past use of a facility within the same ACO.

In the facility choice model, each patient is assigned a choice set $\mathcal{CS}_{imd}$. I begin by restricting this set to facilities located within 75 miles of patient i's residential zip code.[23] To account for clinical relevance, I further limit the choice set to facilities that have ever admitted a patient with a diagnosis in disease category d at any point in the claims data. Patients enrolled in Medicare Advantage face network restrictions; to incorporate this into the model, I restrict the choice set of Medicare Advantage enrollees to facilities that are included in the network of the firm providing insurance benefits. For patients enrolled in Traditional Medicare, I do not impose choice set restrictions at the insurer level, as patients enrolled in Traditional Medicare do not face network restrictions.

Likelihood Function

Having described the empirical specification for the demand and supply model, I now turn to the estimation of the facility choice model. I estimate the parameters of the demand model following the two stage approach of Goolsbee and Petrin (2004).

First, I rewrite the utility function in equation 2 by absorbing all variables that vary at the facility-diagnosis-insurer into a mean utility term δ_{jmd}. The rewritten utility function is

$$u_{ijmd} = \delta_{jmd} + z_{ijmd}\gamma + \epsilon_{ijmd} \tag{25}$$

$$\delta_{jmd} = -\alpha_m p_{jd}(r_d, x_j) + \beta q_j + \xi_{jmd} \tag{26}$$

The parameters to estimate in the first step are the vector of preferences for characterisitics that vary at the individual-facility level (e.g. travel distance) γ and the facility-diagnosis-insurer-year fixed effects $\delta = \{\delta_{jmd}\}_{\forall j,m,d}$. I estimate these parameters using the claims-level data in a maximum likelihood estimation routine. The log-likelihood function is

$$logL(\delta, \gamma) = \sum_{i,m,d} \sum_{j \in \mathcal{CS}_{imd}} h_{ijmd} \times log(\frac{exp\{\delta_{jmd} + z_{ijmd}\gamma\}}{\sum_{j' \in \mathcal{CS}_{imd}} exp\{\delta_{j'md} + z_{ij'md}\gamma\}}) \tag{27}$$

where $\frac{exp\{\delta_{jmd}+z_{ijmd}\gamma\}}{\sum_{j' \in \mathcal{CS}_{imd}} exp\{\delta_{j'md}+z_{ij'md}\gamma\}}$ is the probability that patient i with diagnosis category d covered by insurer m at time t chooses provider j.

Maximizing the log-likelihood function over the space of (γ, δ) is computationally burdensome due to the very high dimensionality of δ. To reduce the computational burden, I follow the approach of Goolsbee and Petrin (2004) by concentrating out the likelihood function and

[23] Only 0.9% of observed visits fall outside this radius, which are dropped from the sample.

maximize only over γ. Let $\tilde{\gamma}$ be a candidate value for the vector of parameters that govern individual preferences. For a given value of $\tilde{\gamma}$, I use the Berry (1994) inversion with the Berry et al. (1995) contraction mapping to compute the unique set of facility mean utilities $\delta(\tilde{\gamma})$ that match the model predicted market shares of facility j for diagnosis category d for insurer category m with the observed market shares.[24]

Define $\mathcal{I}_{md}$ as the set of patients who are in diagnosis category d covered by insurer m at time t, N_{md} as the number of patients in the set $\mathcal{I}_{md}$, $\mathcal{J}_{md}$ as the set of facilities who treat patients of insurer category m and diagnosis category d, and s^{Data}_{jmd} as the facility-insurer-diagnosis specific market shares observed in the data. The market share matching condition is

$$s^{\text{Data}}_{jmd} = \frac{1}{N_{md}} \sum_{i \in \mathcal{I}_{md}} \frac{exp\{\delta_{jmd} + z_{ijmd}\gamma\}}{\sum_{j' \in \mathcal{CS}_{imd}} exp\{\delta_{j'md} + z_{ij'md}\gamma\}} \ \forall j \in \mathcal{J}_{md}, \forall m, d. \tag{28}$$

In practice, for a candidate value of $\tilde{\gamma}$ and a given market at the insurer-diagnosis level, the mean utility terms are solved for by iterating over

$$\delta^{l+1}_{jmd} = \delta^{l}_{jmd} + log(s^{\text{Data}}_{jmd}) - log(\frac{1}{N_{md}} \sum_{i \in \mathcal{I}_{md}} \frac{exp\{\delta^{l}_{jmd} + z_{ijmd}\gamma\}}{\sum_{j' \in \mathcal{CS}_{imd}} exp\{\delta^{l}_{j'md} + z_{ij'md}\gamma\}}) \ \forall j \in \mathcal{J}_{md} \tag{29}$$

where l is the iteration stage. This iterative procedure recovers each mean utility $\delta(\tilde{\gamma})$ for a guess of $\tilde{\gamma}$. Thus, the optimization problem to estimate the individual preferences γ and mean utility preferences δ reduces to just a search over γ,

$$\max_{\gamma} logL(\gamma, \delta(\gamma)) = \sum_{i,m,d} \sum_{j \in \mathcal{CS}_{imd}} h_{ijmd} \times log(\frac{exp\{\delta_{jmd} + z_{ijmd}\gamma\}}{\sum_{j' \in \mathcal{CS}_{imd}} exp\{\delta_{j'md} + z_{ij'md}\gamma\}}) \tag{30}$$

$$\text{s.t. } s^{\text{Data}}_{jmd} = \frac{1}{N_{md}} \sum_{i \in \mathcal{I}_{md}} \frac{exp\{\delta_{jmd} + z_{ijmd}\gamma\}}{\sum_{j' \in \mathcal{CS}_{imd}} exp\{\delta_{j'md} + z_{ij'md}\gamma\}} \ \forall j \in \mathcal{J}_{md}, \forall m, d. \tag{31}$$

After the mean utility values are recovered from the first stage, I estimate the parameters absorbed by the mean utilities that govern price and quality sensitivity by estimating the parameters of the regression in equation 26, leaving ξ_{jmd} as a structural error term. Firms' knowledge of ξ_{jmd} renders $p_{jd}(x_j, r_d)$ and q_j as endogenous. To address the endogeneity of spending and quality, I estimate α_m, β using two-stage least squares. I construct instruments

[24] The observed market shares are constructed using the universe of outpatient claims.

using cost shifters that affect price and quality, but are not correlated ξ_{jmd}. The rationale behind these instruments is that cost shifters are correlated with spending and quality, but are uncorrelated with unobserved consumer preferences. The cost shifters I use are a housing price index, physician wages, physician assistant wages, nurse practitioner wages and a cost index constructed by CMS.

Identification

Identification of the facility choice model relies on cross-sectional and longitudinal variation in choice sets, in addition to differences in facility and patient characteristics. The vector of parameters γ are identified from how patients' choice probabilities respond to variation in characteristics included in z_{ijmd} across the set of facilities available to them within the same diagnosis-insurer-year. For example, the coefficient associated with the distance between the patient and facilities is identified by covariation between choice probabilities and distance. Identification of mean utility values δ_{jmd} are identified from the market share matching constraint, as in Berry et al. (1995).[25] Identification of the spending and quality sensitivity parameters relies on appropriate exclusion restrictions, specifically the excluded cost shifters.

4.2 Supply Estimation

Taking estimates for demand parameters as given, I discuss the empirical specification and estimation strategy for estimating supply parameters.

Empirical Specification

As described in the supply model, facilities who participate in an ACO face a sharing rule that dictates how the ACO's performance is split across participants. Let λ_k be the proportion of ACO performance kept by the ACO, and $1 - \lambda_k$ is the proportion of ACO performance kept by the government. I specify the empirical sharing rule as

$$w_f = \begin{cases} 0 \text{ if } k = 0 \\ \lambda_k \text{ if } k \neq 0, \mathcal{J}_f = \mathcal{J}_k \\ \lambda_k \frac{W_f}{\sum_{f' \in \mathcal{J}_k} W_{f'}} \text{ if } k \neq 0, \mathcal{J}_f \neq \mathcal{J}_k \end{cases} \tag{32}$$

[25]Because mean utilities are only identified up to an additive constant, I normalize the mean utility of the outside option to zero within each diagnosis–insurance category.

where $\mathcal{J}_k$ is the set of facilities that participate in ACO k and $\mathcal{J}_f$ is the set of facilities in system f. It is worth noting that the set of facilites in system f is a subset of the set of facilities in ACO k, $\mathcal{J}_f \subseteq \mathcal{J}_k$, where $\mathcal{J}_f = \mathcal{J}_k$ if system f is the sole participant in ACO k and $\mathcal{J}_f \subset \mathcal{J}_k$ if multiple systems participate in ACO k. If hospital system f does not participate in an ACO, as indicated by $k = 0$, w_f takes on the value 0. If hospital system f is the only participant in ACO k, that is $\mathcal{J}_f = \mathcal{J}_k$, then hospital system f internalizes the entire performance of ACO and keeps the performance leftover after the government keeps its share. If there are multiple systems that participate in ACO k, $\mathcal{J}_f \neq \mathcal{J}_k$, then the systems split the performance according to the sharing rule $\lambda_k \times \frac{W_f}{\sum_{f' \in \mathcal{J}_k W_{f'}}}$, where W_f is an observable measure that captures the division of ACO performance. For the empirical specification, I use the aggregate spending of firm f across all patients in the year prior to forming the ACO as W_f, which captures the notion that bigger systems will have keep larger share of the ACO's performance.

Next, I discuss the empirical specification of the marginal cost function. Building off of equation 10, I decompose the non-quality related costs c_j^o into a matrix of exogenous observable characteristics Z_j, a vector of parameters ψ_0 that govern the shape of the non-quality component of the cost curve and a structural error term ε_j. The non-quality related costs are

$$c_j^o = Z_j \psi_0 + \varepsilon_j. \tag{33}$$

Next, I decompose the slope of marginal cost with respect to quality into a matrix of exogenous observable characteristics Z_j, a vector of parameters ψ_1 that govern the shape of the quality related component of the cost curve and a structural error term η_j.[26] The empirical specification for the slope of the marginal cost curve with respect to quality is

$$\frac{\partial mc_j}{\partial q_j} = Z_j \psi_1 + \eta_j. \tag{34}$$

Finally, I assume that ε_j and η_j are observed by the players, but not observed by the econometrician.

[26]With slight abuse of notation, I write the set of explanatory variables in equation 33 as the same set of explanatory variables in equation 34. These two sets of explanatory variables need not be the same, though I use the same set of explanatory variables in practice.

Recovering Marginal Cost Terms

Prior to discussing my estimation strategy for $\{\psi_0, \psi_1\}$, I discuss the procedure to recover the marginal cost terms for all facilities using systems' optimality conditions. Given observed quantity of services, quality of care and estimates of the demand system, I recover the unobserved marginal costs mc_j and each facilities' slope of marginal cost with respect to quality $\frac{\partial mc_j}{\partial q_j}$ by inverting the systems' first-order conditions. Specifically, I assume that the quantity of services and quality of care that are observed in the data are the result of firms maximizing profits. Under this assumption, I solve for the vector of marginal costs that would rationalize the equilibrium outcomes of quantity of services and quality of care, characterized by equations 21 and 22.

First, I solve for the vector of marginal costs by rearranging equation 21, following Berry et al. (1995). The vector of facility marginal costs are solved for as

$$\boldsymbol{mc} = \boldsymbol{p(x)} + (\boldsymbol{\Delta^x})^{-1} diag(\frac{\boldsymbol{\partial p}}{\boldsymbol{\partial x}})\boldsymbol{D} + (\boldsymbol{\Delta^x})^{-1} diag(\boldsymbol{w})\frac{\boldsymbol{\partial \pi}}{\boldsymbol{\partial x}} \tag{35}$$

where the terms $\boldsymbol{p(x)}$, $\frac{\boldsymbol{\partial p}}{\boldsymbol{\partial x}}$ and $\boldsymbol{w}$ are observed in the data, and the terms $\boldsymbol{\Delta^x}$, $\boldsymbol{D}$ and $\frac{\boldsymbol{\partial \pi}}{\boldsymbol{\partial x}}$ are estimated from the demand system.

With the marginal costs of each facility recovered, I solve for the vector of each facility's slope of marginal cost with respect quality using the system of equations generated by the first order conditions with respect to quality of care characterized in equation 22, combined with the recovered levels of marginal costs from equation 35. The vector of each facilities' slope of marginal cost with respect to quality is

$$\frac{\boldsymbol{\partial mc}}{\boldsymbol{\partial q}} = diag(\boldsymbol{D})^{-1}\big(\boldsymbol{\Delta^q}(\boldsymbol{p(x)} - \boldsymbol{mc}) + diag(\boldsymbol{w})\frac{\boldsymbol{\partial \pi}}{\boldsymbol{\partial q}}\big) \tag{36}$$

where the terms $\boldsymbol{p(x)}$ and $\boldsymbol{w}$ are observed in the data, the terms $\boldsymbol{\Delta^q}$, $\boldsymbol{D}$ and $\frac{\boldsymbol{\partial \pi}}{\boldsymbol{\partial q}}$ are estimated from the demand system, and $\boldsymbol{mc}$ is solved for in equation 35.

With the levels of marginal costs recovered in equation 35 and the slope of marginal costs with respect to quality recovered in 36, I solve for the vector of each facilities' non-quality related component of marginal costs using the specification in equation 10, following Crawford et al. (2019)

$$c_j^o = mc_j - \frac{\partial mc_j}{\partial q_j} q_j, \forall j \tag{37}$$

where each facilities' value of mc_j are solved for in equation 35, each facilities' value of $\frac{\partial mc_j}{\partial q_j}$ are solved for in equation 36 and each facilities' value of q_j are observed in the data.

Estimating Cost Parameters

Using the non-quality related costs for each facility that rationalize the equilibrium, recovered from equations 35, 36 and 37, I estimate ψ_0 by regressing the recovered values of non-quality related marginal costs from equation 37 on the empirical specification characterized in equation 33 using OLS, where the left hand side of the regression are the recovered levels of non-quality related marginal costs. Similarly, I estimate ψ_1 by regressing the values recovered in equation 36 on the empirical specification characterized in equation 34 using OLS.[27]

5 Estimation Results

5.1 Demand Estimates

Table 4 presents estimates of the demand parameters.

Price Sensitivity I find that consumers dislike facilities with higher spending. Consumers who are enrolled in Traditional Medicare without Medicaid coverage have a lower price sensitivity compared to those who are enrolled in Medicare Advantage without Medicaid coverage. This is not surprising, as Traditional Medicare has more generous insurance coverage (more generous coinsurance/copayments) compared to Medicare Advantage.[28] The higher price elasticity of Medicare Advantage patients compared to patients with Traditional Medicare is consistent with studies that attempt to measure price elasticity for Medicare patients who do not supplemental insurance, as in Hoffman (2015).

Quality As shown by the positive coefficient on quality, consumers prefer facilities with higher quality. This is consistent with the literature on patient preferences over facilities with

[27]Although firms have knowledge of their values of the structural error terms ε_j and η_j, I assume the error terms are uncorrelated with cost shifters Z_j.

[28]Patients who are enrolled in Traditional Medicare without Medicaid coverage on average have lower copayments/coinsurance rate than Medicare Advantage, due to supplemental insurances such as Medigap and employer-sponsored supplemental insurance.

	Est.	SE.
Spending ($1,000's)		
TM, No Medicaid	-3.07	(0.08)
MA, No Medicaid	-2.87	(0.09)
Quality	4.61	(0.34)
Distance	-17.83	(0.000)
Distance squared	4.36	(0.000)
Past use hospital	7.25	(0.000)
Past use system	7.96	(0.001)
Past use ACO	9.05	(0.000)
N = 16,212,409		

Table 4: Demand Parameter Estimates

Notes: Parameter estimates are reported with standard errors in parentheses. Distance terms are measured in 100's of miles. Past use variables are a dummy variable that takes value 1 if the patient visited a hospital/system/ACO at a point prior to the current visit.

differentiating health outcome quality measures (Tay, 2003; Gaynor et al., 2016). Traditional Medicare patients are willing to pay about $45 for a one standard deviation increase in quality, compared to about $48 for Medicare Advantage patients.

Distance Consistent with the literature on hospital choice, patients dislike travel distance as evidenced by the negative coefficient on the first order distance term (Gowrisankaran et al., 2015; Kessler and McClellan, 2000). As distance increases, the marginal disutility from travel distance decreases, as seen by the positive coefficient on the second order distance term.[29] The parameter estimates imply that patients with Traditional Medicare are willing to pay $58, and Medicare Advantage patients $62, to reduce travel distance by one mile.

Past Facility Use Examining past use of a facility, I find that patients are more likely to visit a facility if the facility was previously visited, as evidenced by a positive coefficient, which is consistent with Shepard (2022); Prager (2020). Conditional on other covariates, the probability that a patient chooses a previously visited facility is approximately 4.5 times higher than choosing a facility with no prior visit.

Turning to the coefficient on past use of a facility in the same hospital system, I find

[29]The inflection point at which the marginal utility from travel distance occurs at approximately 204 miles ($-17.83 + 2 \times 4.36 \times dist = 0$), which means the marginal utility of distance is always negative within the 75 mile choice set imposed in the empirical specification.

	Est.	SE.
Non-quality costs (ψ_0)		
Constant	352	(67.22)
Quality costs (ψ_1)		
Constant	1,318	(45.94)
Average Marginal Cost	$1,617	
Average Price-Cost Margin	$181	
Average Markup	14.92%	
N = 678 (Facility-Years)		

Table 5: Cost Parameter Estimates

Notes: Parameter estimates are reported with standard errors in parentheses. Model-implied averages are calculated at the estimated parameter values and data points.

that patients who have previously visited a facility within a system are more likely to choose another facility in the same system in the future. Conditional on other covariates, the probability that a patient chooses a facility after previously visiting another facility within the same system is approximately 4.7 times higher than choosing a facility in a different system. Likewise, patients who have previously visited a facility with an ACO are more likely to choose another facility in the same ACO. Conditional on other covariates, the probability that a patient chooses a facility after previously visiting another facility within the same ACO is approximately 4.7 times higher than choosing a facility in a different ACO. This finding is consistent with the results of Chernew et al. (2021), who find that referral behavior has a strong effect on patients' choice for a provider, even offsetting disutilities from travel distance and out of pocket costs. Similarly, Ho and Pakes (2014) find that providers are more likely to refer patients to another provider in the same ACO. In addition, Baker et al. (2016) find that system ownership of healthcare providers increases the likelihood that a patient selects another provider within the same system, even if that provider is high-cost and lower-quality.

5.2 Supply Estimates

Table 5 presents parameter estimates of the supply model.

Examining Table 5, non-quality related costs are approximately $352 on average, while the average cost of increasing quality by 1 unit is $1,318. Together, the average cost of treating one patient is approximately $1,617. Examining per patient markups, the average price-cost margin per patient is approximately \$181, which translates to average markup is 15%.[30] The estimated markups are consistent with Medicare cost reports for outpatient hospital based care (CMS, 2025b).

6 Counterfactuals

In this section, I perform counterfactual analyzes to quantify the welfare impact of ACOs. Prior to discussing counterfactual experiments, I construct measures for consumer welfare, producer welfare and government spending.

Consumer Welfare I define the aggregate consumer surplus across all individuals as the ex-ante expected utility prior to realizing idiosyncratic shocks ϵ_{ijmd}, following Small and Rosen (1981).[31] The aggregate consumer surplus in the market is

$$CW = \sum_i \frac{1}{|\alpha_{m(i)}|} log(\sum_{j \in CS_{imd}} exp\{-\alpha_m p_{jd}(r_d, x_j) + \beta q_j + z_{ijmd}\gamma + \xi_{jmd}\}). \tag{38}$$

Producer Welfare For firms, the aggregate welfare outcome is the profit aggregated across all systems in the market

$$PW = \sum_f \Pi_f. \tag{39}$$

Government Spending As ACO incentives change in the counterfactual scenarios, the calculation of government spending is calculated as the sum of total reimbursements, combining both outpatient reimbursements and inpatient reimbursements, and net payout for ACO performance

$$GW = \sum_j p_j(x_j) D_j(\boldsymbol{x}, \boldsymbol{q}) + \sum_j (1 - q_j) Y_j D_j(\boldsymbol{x}, \boldsymbol{q}) + \sum_k (1 - \lambda_k) * \pi_k(\boldsymbol{x}, \boldsymbol{q}) \tag{40}$$

[30]For facilities that do not participate in ACOs, I calculate the price-cost margin as the per-patient payment minus the total marginal cost of treating one patient. For facilities that participate in an ACO, the price-cost margin adds an additional term for the performance of the ACO.

[31]The ex-ante expected utility has a closed form solution as a result of the T1EV assumption on ϵ_{ijmd}. I exclude Euler's constant from the expression for ex-ante patient utility.

where $\pi_k(\boldsymbol{x}, \boldsymbol{q})$ may be positive or negative. If $\pi_k(\boldsymbol{x}, \boldsymbol{q})$ is positive, the government gives a payment to the ACO captured by $(1 - \lambda_k) * \pi_k(\boldsymbol{x}, \boldsymbol{q})$; otherwise, the government penalizes the ACO for negative performance and receives a payment of $(1 - \lambda_k) * \pi_k(\boldsymbol{x}, \boldsymbol{q})$.

6.1 Removing ACOs

Using the model, estimates of the demand system and recovered marginal costs, I simulate a counterfactual where I remove the incentives of ACOs from the supply side by setting w_f equal to 0 for each systems. In addition, I remove the effects of steering patients through ACOs on the demand side. I compare the equilibrium outcomes and welfare measures of this counterfactual to the equilibrium observed in the data to evaluate and quantify the welfare impacts of ACOs. Table 6 presents the statistics of the endogenous outcomes in this simulated equilbirum, including outpatient payments and quality, in addition to the aggregate welfare outcomes of total consumer surplus, total insurer surplus, and government spending.

Table 6: Impact of ACOs

	Observed ACO Affiliation	Without ACOs	Impact
Facility-Level Outcomes			
Outpatient payment ($p_j(x_j)$)	\$1,826	\$1,917	-\$91
Quality (q_j)	0.96	0.94	0.02
Aggregate Welfare			
Total consumer surplus	2.01	1.76	0.25
TM Consumers	1.37	1.20	0.17
MA Consumers	0.64	0.56	0.08
Total system surplus	1.74	1.63	0.11
ACO Participants	0.82	0.67	0.15
Non-Participants	0.92	0.96	-0.04
Government Spending	12.76	13.49	-0.73

Note: Facility level outcomes are the averages of the endogenous variables. Aggregate outcomes are measured in billions of dollars. Counterfactual analysis performed on the year 2019. ACO participants refers to systems that participated in an ACO prior to removing ACO incentives. Non-participants refers to systems that did not participate in an ACO prior to removing ACO incentives.

Facility Outcomes Examining facility level outcomes, implementing ACO incentives from the market decreases the average outpatient payment per patient by \$91 (-4.74%). With patients facing an average of approximately 6 outpatient visits per year, this results in a decrease of approximately $546 in annual outpatient spending per patient. The average quality of care increased by 0.02 units (2.13%), which can be interpreted as an expected decrease of $260 of annual inpatient spending per patient. Quantitatively, this implies that implementing ACO incentives in the market decreases annual per-patient spending by an average of approximately $806.

Turning to the mechanisms that led to this result, systems who participate in an ACO have an increased the marginal disutility facilities receive from increasing quantity of services, as well as increased marginal benefit facilities receive from increasing quality of care. 31% of systems that participated in an ACO faced a spending-quality tradeoff, which is consistent with Reddig (2024). Of these systems, 73% decrease outpatient spending, while also reducing quality of care. Another mechanism in play are competition effects onto systems that do not participate in ACOs; ACO participants' more competitive choices of quantity of services and quality of care incentivize systems that do not participate in an ACO to increase their quality of care and decrease their quantity of services, to capture demand.

Consumer Welfare Implementing ACOs in the market increases total consumer surplus by $0.25 billion (14%), driven by the decrease in outpatient spending and the increase in quality of care. Both Traditional Medicare and Medicare Advantage patients are positively affected by removing ACOs, even though ACO incentives primarily depend on the spending of Traditional Medicare patients. The welfare gains from ACOs to Traditional Medicare patients are $0.17 billion, while the spillover effect onto Medicare Advantage patients' consumer welfare is $0.08 billion. Welfare gains to Medicare Advantage patients account for roughly 32% of the total consumer welfare gains, implying that ignoring these spillover effects biases the welfare evaluation of the ACO program.

Producer Welfare Moving onto producer welfare, implementing ACOs increases profits by $0.11 billion (6.13%). Systems that participated in an ACO faced a $0.15 billion gain in profit, while Non-Participants faced a $0.04 billion dollar loss in profits. Examining systems that participate in an ACO, ACO participants received less revenue and higher costs per consumer due to lower outpatient spending and higher quality, but had higher demand for the same reasons. In addition, these systems are able to leverage the "network" formed by ACOs to steer patients within the ACO, effectively increasing demand. Finally, participants in an ACO receive a split of the performance of the ACO. On average, systems that participate saved $770 per patient, relative to the benchmark, effectively increasing profit for systems that participate in an ACO. Turning to systems that did not participate in an ACO, these systems faced lower profit, due to lower outpatient revenue and higher costs due to higher quality. While these two effects would increase demand, non-participating systems are not able to steer patients through the same mechanisms as ACOs, which lowers demand for non-participating systems relative to ACO participants. In addition, non-participating systems do not receive a financial reward for decreasing patient spending, which further reduces profit compared to ACO participants

Government Spending Examining government expenditure, implementing ACO incentives decreases government expenditure by 0.73 billion (-5.41%). This is driven mostly by lower aggregate outpatient and inpatient spending, though the government does need to provide systems that participate in an ACO the benchmark per assigned patient.

Total Welfare Overall, ACOs are welfare improving, amounting to a total welfare gain of $1.09 billion, though there are heterogeneous impacts to firms. Traditional Medicare and Medicare Advantage patients are better off in the regime with ACOs. ACO participants receive higher profit under the regime with ACOs, while non-participants are worse off. Government spending decreases under the equilibrium with ACO incentives.

7 Conclusion

This paper develops and estimates a structural model of demand and supply in the Medicare hospital outpatient services market to evaluate the welfare impacts of Accountable Care Organizations (ACOs). The model incorporates facilities' endogenous choices of both service quantity and quality of care. Participation in an ACO incentivizes systems to reduce the quantity of services while improving quality, which can lead to either quality overprovision with high spending or quality underprovision with low spending and low quality.

I find that the ACO program is welfare improving. On average, participating systems lower outpatient spending while raising quality of care, and non-participants respond competitively by doing the same. These adjustments increase consumer welfare and reduce government spending.

My analysis focuses on the equilibrium outcomes of competition over service quantity and care quality, treating ACO participation as exogenous. In reality, participation in an ACO is an endogenous choice that I abstract away from, which may bias welfare estimates by ignoring the fixed costs of ACO formation and their potential correlation with marginal costs. The decision to form or join an ACO is complex, as systems can either establish a stand-alone ACO or coordinate with other systems to form one jointly. This choice is further complicated by the need for mutual agreement among participating systems. Future work that endogenizes the joint decision of ACO formation across systems would be a promising direction for future research.

References

Aswani, A., Z.-J. M. Shen, and A. Siddiq (2019). Data-driven incentive design in the medicare shared savings program. *Operations Research 67*(4), 1002–1026.

Baker, L. C., M. K. Bundorf, and D. P. Kessler (2016). The effect of hospital/physician integration on hospital choice. *Journal of Health Economics 50*(C), 1–8.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics 25*(2), 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Chernew, M., Z. Cooper, E. L. Hallock, and F. Scott Morton (2021). Physician agency, consumerism, and the consumption of lower-limb mri scans. *Journal of Health Economics 76*, 102427.

CMS (2025a). Aco reach performance year 2023 financial and quality results fact sheet.

CMS (2025b). Cost reports – HCRIS Data on Medicare and Other Providers. `https://www.cms.gov/data-research/statistics-trends-and-reports/cost-reports`. Accessed September 30, 2025.

CMS (2025c). National health expenditure (nhe) fact sheet.

Cooper, Z., S. V. Craig, M. Gaynor, and J. Van Reenen (2019). The price ain't right? hospital prices and health spending on the privately insured. *Quarterly Journal of Economics 134*(1), 51–107.

Crawford, G. S., O. Shcherbakov, and M. Shum (2019). Quality overprovision in cable television markets. *American Economic Review 109*(3), 956–995.

Cutler, D. M. (1995). The incidence of adverse medical outcomes under prospective payment. *Econometrica 63*(1), 29–50.

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

Einav, L., A. Finkelstein, and N. Mahoney (2025). Producing health: Measuring value added of nursing homes. *Econometrica 93*(4), 1225–1264.

Ellis, R. P. and T. G. McGuire (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics*.

Fan, Y. (2013, August). Ownership consolidation and product characteristics: A study of the us daily newspaper market. *American Economic Review 103*(5), 1598–1628.

Fudenberg, D. and J. Tirole (1991). *Game Theory*. Cambridge, MA: MIT Press.

Gaynor, M., J. Li, and W. B. Vogt (2004). Physician incentives in health maintenance organizations. *Journal of Political Economy 112*(4), 915–931.

Gaynor, M., C. Propper, and S. Seiler (2016, November). Free to choose? reform, choice, and consideration sets in the english national health service. *American Economic Review 106*(11), 3521–57.

Gaynor, M. and W. B. Vogt (2003). Competition among hospitals. *The RAND Journal of Economics 34*(4), 764–785.

Geweke, J., G. Gowrisankaran, and R. J. Town (2003). Bayesian inference for hospital quality in a selection model. *Econometrica 71*(4), 1215–1238.

Gluzman, J. and Y. Zhou (2025). Regulation of contract generosity in selection markets: The case of medicare advantage. Technical report, Working Paper.

Goolsbee, A. and A. Petrin (2004). The consumer gains from direct broadcast satellites and the competition with cable tv. *Econometrica 72*(2), 351–381.

Gowrisankaran, G., A. Nevo, and R. Town (2015, January). Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review 105*(1), 172–203.

Gowrisankaran, G. and R. J. Town (1999). Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics 18*(6), 747–767.

Hackmann, M. B. (2019). Incentivizing better quality of care: The role of medicaid and competition in the nursing home industry. *American Economic Review 109*(5), 1684–1716.

Ho, K. and R. S. Lee (2017). Insurer competition in health care markets. *Econometrica 85*(2), 379–417.

Ho, K. and A. Pakes (2014, December). Hospital choices, hospital prices, and financial incentives to physicians. *American Economic Review 104*(12), 3841–84.

Hoffman, G. (2015, February). Cost-sharing, physician utilization, and adverse selection among medicare beneficiaries with chronic health conditions. *Medical Care Research and Review 72*(1), 49–70. Epub 2014 Dec 23.

Kessler, D. P. and M. B. McClellan (2000, May). Is hospital competition socially wasteful? *The Quarterly Journal of Economics 115*(2), 577–615.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

Prager, E. (2020, October). Healthcare demand under simple prices: Evidence from tiered hospital networks. *American Economic Journal: Applied Economics 12*(4), 196–223.

Reddig, K. (2024, October). Designing physician incentives and the cost-quality tradeoff: Evidence from accountable care organizations. Working Paper.

Shepard, M. (2022). Hospital network competition and adverse selection: Evidence from the massachusetts health insurance exchange. *American Economic Review 112*(2), 578–615. NBER Working Paper No. 22600.

Small, K. A. and H. S. Rosen (1981). Applied welfare economics with discrete choice models. *Econometrica 49*(1), 105–130.

Tay, A. (2003). Assessing competition in hospital care markets: The importance of accounting for quality differentiation. *The RAND Journal of Economics 34*(4), 786–814.

A Appendix

A.1 ACO Assignment Rule

Each year, a patient is assigned to an ACO if the ACO has the highest level of annual primary care patient, relative to all other ACOs. The HCPCS codes that fall under the category of primary care spending are 99201-99215 (New patient & Established patient visits for non Skilled Nursing Facility settings), 99304-99340 (New patient & Established patient visits for Skilled Nursing Facility settings), 99341-99350 (New patient & Established patient visits for home services). Given that my sample consists only of outpatient and inpatient hospital visits, the only relevant HCPCS codes are 99201-99215. Table A.1 provides a description of each relevant HCPCS codes used for patient assignment

Table A.1: Primary Care HCPCS Codes

HCPCS Code	Description
99201	New Patient, brief
99202	New Patient, limited
99203	New Patient, moderate
99204	New Patient, comprehensive
99205	New Patient, extensive
99211	Established Patient, brief
99212	Established Patient, limited
99213	Established Patient, moderate
99214	Established Patient, comprehensive
99215	Established Patient, extensive

A.2 CCS Diagnosis Categories

Table A.2 presents the diagnosis categories used in the empirical analysis. Each diagnosis is placed into one of eighteen mutually exclusive categories, grouped by CCS. Diagnoses that fall into diseases of the circulatory system are the most prevalent, while birth related diagnoses are the two least common.

A.3 Measuring Quantity of Services

Following the literature on payments to facilities (Gaynor and Vogt, 2003; Gowrisankaran et al., 2015; Ho and Lee, 2017), I first define p_{ijdt} as the payment facility j receives for treating patient i with diagnosis category d at time t, which is observable in the data at the claim level. Using CMS reimbursement rate data, I define r_{dt} as the reimbursement rate for diagnosis d at time t, where the reimbursement rate is measured in dollars per unit of services. Let x_{ijt} be the quantity of services provided to patient i by facility j at time t, where x_{ijt} is measured in units of services. I measure quantity of services x_{ijt} (measured in units of services) as the ratio between the payment p_{ijdt} (measured in dollars) and r_{dt} (measured in dollars per unit of service)

Table A.2: Percent by Diagnosis Group

Diagnosis Group	Percent
Diseases of the circulatory system	15.18
Factors influencing health status	13.37
Neoplasms (Cancer, tumours)	12.52
Endocrine; nutritional; and metabolic diseases and immunity disorders	10.87
Diseases of the musculoskeletal system and connective tissue	10.20
Diseases of the genitourinary system	7.29
Diseases of the nervous system and sense organs	6.03
Mental Illness	6.00
Diseases of the respiratory system	4.76
Diseases of the digestive system	3.67
Diseases of the blood and blood-forming organs	2.58
Diseases of the skin and subcutaneous tissue	2.31
Injury and poisoning	2.26
Infectious and parasitic diseases	1.22
Unclassified	1.48
Complications of pregnancy; childbirth; and the puerperium	0.13
Congenital anomalies (birth defects)	0.13

$$x_{ijt} = \frac{p_{ijdt}}{r_{dt}}. \tag{A.1}$$

Following Ho and Lee (2017), to recover a measure for quantity of services at the facility-year pair, I take the average of x_{ijt} over each patient in the facility-year pair

$$x_{jt} = \frac{1}{N_{jt}} \sum_{i \in \mathcal{I}_{jt}} x_{ijt}$$

where $\mathcal{I}_{jt}$ is the set of patients who choose facility j at time t, and N_{jt} is the number of patients who choose facility j at time t.

A.4 Measuring Quality of Care

I use the detailed claims data and health outcomes of transition from outpatient to inpatient within 30 days for discharge to construct a quality measure for each facility-year. While tempting to compute the average health outcome for all patients within a facility-year, this method fails to address observable selection of patients to facilities and unobservable selection, which is well documented in the literature (Gowrisankaran and Town, 1999; Geweke et al., 2003), Geweke et al. (2003). Following the literature, I estimate a linear probability model of the form

$$Outcome_{ijdt} = \rho_{jt} + \vartheta_d + X_i\varphi + \iota_{ijdt} \tag{A.2}$$

where ρ_{jt} is a facility-year fixed effect, ϑ_d is a diagnosis level fixed effect that controls for persistent differences in transition rates into inpatient care for diagnoses, φ is a vector of parameters associated with the matrix of patient-level controls X_i and ι_{ijdt} is an error term. I estimate ρ_{jt}, ϑ_d, φ using OLS.

Following Cooper et al. (2019), I measure the quality at the facility-year level by calculating the fitted value of the regression in equation A.2, evaluated at the mean value of diagnosis prevalence and patient controls, evaluated across the entire sample.

$$q_{jt} = 1 - (\hat{\rho}_{jt} + \hat{\vartheta}_d + \bar{X}_i\hat{\varphi}) \tag{A.3}$$